# THE EVOLUTION OF MUSIC

EDITED BY: Leonid Perlovsky and Aleksey Nikolsky
PUBLISHED IN: Frontiers in Psychology, Frontiers in Neuroscience and Frontiers in Sociology



frontiers Research Topics



frontiers

**Frontiers eBook Copyright Statement**

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-88966-286-9
DOI 10.3389/978-2-88966-286-9

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

# THE EVOLUTION OF MUSIC

Topic Editors:
**Leonid Perlovsky,** Northeastern University, United States
**Aleksey Nikolsky,** Braavo! Enterprises, United States

**Citation:** Perlovsky, L., Nikolsky, A., eds. (2020). The Evolution of Music. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-286-9

# Table of Contents


*05* ***Editorial: The Evolution of Music***
Aleksey Nikolsky and Leonid Perlovsky

*11* ***The Measurement of Aesthetic Emotion in Music***
Leon Crickmore

*15* ***How Music and Instruments Began: A Brief Overview of the Origin and Entire Development of Music, From its Earliest Stages***
Jeremy Montagu

*27* ***Simultaneous Cooperation and Competition in the Evolution of Musical Behavior: Sex-Related Modulations of the Singer's Formant in Human Chorusing***
Peter E. Keller, Rasmus König and Giacomo Novembre

*41* ***Neurodiversity, Giftedness, and Aesthetic Perceptual Judgment of Music in Children With Autism***
Nobuo Masataka

*48* ***What Vowels Can Tell Us About the Evolution of Music***
Gertraud Fenk-Oczlon

*54* ***White Matter Correlates of Musical Anhedonia: Implications for Evolution of Music***
Psyche Loui, Sean Patterson, Matthew E. Sachs, Yvonne Leung, Tima Zeng and Emily Przysinda

*64* ***Evolutionary Musicology Meets Embodied Cognition: Biocultural Coevolution and the Enactive Origins of Human Musicality***
Dylan van der Schyff and Andrea Schiavio

*82* ***The Role of the Baldwin Effect in the Evolution of Human Musicality***
Piotr Podlipniak

*94* ***A Joint Prosodic Origin of Language and Music***
Steven Brown

*114* ***The Paradox of Isochrony in the Evolution of Human Rhythm***
Andrea Ravignani and Guy Madison

*127* ***The Evolution of Musicality: What Can Be Learned From Language Evolution Research?***
Andrea Ravignani, Bill Thompson and Piera Filippi

*134* ***Commentary: The 'Musilanguage' Model of Language Evolution***
Aleksey Nikolsky

*138* ***"The Two Brothers": Reconciling Perceptual-Cognitive and Statistical Models of Musical Evolution***
Steven Jan

*153* ***Computational Approach to Musical Consonance and Dissonance***
Lluis L. Trulla, Nicola Di Stefano and Alessandro Giuliani

*164* ***Music Evolution in the Laboratory: Cultural Transmission Meets Neurophysiology***
Massimo Lumaca, Andrea Ravignani and Giosuè Baggio

175 ***Music and the Meeting of Human Minds***
Alan R. Harvey

181 ***Rhesus Monkeys (*Macaca mulatta*) Sense Isochrony in Rhythm, but Not the Beat: Additional Support for the Gradual Audiomotor Evolution Hypothesis***
Henkjan Honing, Fleur L. Bouwer, Luis Prado and Hugo Merchant

196 ***The Human Nature of Music***
Stephen Malloch and Colwyn Trevarthen

217 ***The Overlooked Tradition of "Personal Music" and its Place in the Evolution of Music***
Aleksey Nikolsky, Eduard Alekseyev, Ivan Alekseev and Varvara Dyakonova

260 ***The Pastoral Origin of Semiotically Functional Tonal Organization of Music***
Aleksey Nikolsky

frontiers in Psychology

EDITORIAL
published: 28 October 2020
doi: 10.3389/fpsyg.2020.595517

# Editorial: The Evolution of Music

*Aleksey Nikolsky*[1]* *and Leonid Perlovsky*[2]

[1] Braavo Enterprises, Los Angeles, CA, United States, [2] Department of Psychology, Northeastern University, Boston, MA, United States

**Keywords: interdisciplinary approach, scientific/humanitarian methods, ethnomusicology, empirical musicology, musilanguage, musicological analysis, organology, geomusicology**

## Editorial on the Research Topic

## The Evolution of Music

Two decades ago, Wallin et al. (1999) opened a new page in the quest for the origins of music by integrating the methodologies and data coming from numerous disciplines:

- physical anthropology
- paleoneurobiology
- ethology
- biocultural evolution
- systematic musicology
- semiotics
- historical linguistics
- developmental psychology

This volume updates this multi-disciplinary approach by further advancing the fields defined by Wallin et al. and by introducing and specifying new fields of inquiry:

- organology
- musicological analysis
- geomusicology
- demography
- information theory
- statistic modeling
- (paleo)aesthetics
- (paleo)phonology
- biological motion

In addition, this volume addresses a number of interdisciplinary problems that were identified by the contributing authors[1]. The latter issue has recently become critical: the very idea of multidisciplinary study of music has been questioned. There is a growing conviction amongst Western scholars of ethnomusicological background that humanities and sciences are fundamentally split, and the scientific approach somehow introduces an "anti-humanitarian" bias (Parncutt, 2017). According to this view, specialists in sciences should adjust their methodologies to comply with the conventions of political correctness currently adopted by many Western specialists

[1]This includes not only those papers that were submitted and published as part of our Research Topic, but also the ones that were not published (for various reasons).

OPEN ACCESS

***Edited and reviewed by:***
*Bernhard Hommel,*
*Leiden University, Netherlands*

***Correspondence:**
*Aleksey Nikolsky*
*aleksey@braavo.org*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *16 August 2020*
***Accepted:*** *18 September 2020*
***Published:*** *28 October 2020*

***Citation:***
*Nikolsky A and Perlovsky L (2020) Editorial: The Evolution of Music. Front. Psychol. 11:595517. doi: 10.3389/fpsyg.2020.595517*

in musicology and social sciences. This view exploits the argument that the public trust in science supports the "scientific hegemony" in an ongoing cultural "warfare" between disciplines of art and science, thereby precluding the advance in human knowledge (Cohen, 2001). This argument was introduced half-a-century ago in a popular book by Snow (1964). Despite being convincingly debunked by Wilson (1998) and Gould (2003), it resurfaced again in the ideology of "new mysterianism," propagated by Chomsky (2016) and McGinn (2015). Their intellectual weight has made the call for "humanizing" science more appealing to scientists.

Mysterianist and anti-scientific sentiments found fertile ground amongst many musicians and musicologists that regard a scientific approach as being "bound to the science lab" and therefore irrelevant to music practice—even detrimental to the expressive efficacy of music-makers (Woody, 2004). Indeed, differences in jargon between musicological[2] and psychoacoustic disciplines set a barrier against co-understanding, which preserves the popular myths about perception/production of music (Juslin et al., 2012). Many believe that music is incommensurable, irrational, and mystic—unsuitable for objective investigation (Dubal, 1985) and demanding an intuitive approach (Woody, 2000).

A big role in the spread of this attitude has been played by the ongoing trend of "scientism" in influential Western schools of composition, and the critical acclaim they have been receiving from well-established musical critics and musicologists (Regelski, 2014). Overwhelmingly, the music works composed after WWII and esteemed by academia bear a strong flavor of scholasticism which resembles that of religious scholarly traditions of pre-industrial cultures. The New Music of the West has deliberately exploited the "scientific" image in employing an invention-like approach, equating the method of strictly following a set of newly invented abstract compositional principles to mathematics (Babbitt, 2003). Such music usually violates the psycho-physiological restraints (Thomson, 2010), contributing to the public impression of its fundamental indigestibility (Lerdahl, 1992).

However, in toto, the situation is far from a global scientific/humanitarian schism: science does not defy humanities—it only corrects erroneous traditional beliefs. The only humanitarian discipline that consistently defies scientific methodology is Western ethnomusicology. In the past 40 years it has adopted the view that there is no such thing as "music" but myriad of "musics" (Becker, 1986), each requiring its unique frame of investigation (Nettl, 2010). The very application of a scientific method is viewed here as exercising Eurocentric political power over non-European cultures (Messner, 1993)—seen even in merely calling a non-European artifact "music" (Bohlman, 1999). This highly politicized philosophy resulted in abandoning comparative studies of music systems (Savage and Brown, 2013), of the origins and evolution of music (Nettl, 2005), and of music analysis (Nattiez, 2012), especially pronounced in the US and UK ethnomusicological schools (Zemtsovsky, 2002).

General shift of Western ethnomusicology away from comparative musicology to fractured sociomusicology of isolated musical communities was inspired by concerns for compensating for the earlier Eurocentric bias in ethnographic research prior to the mid-twentieth century. Late twentieth century ethnomusicologists avoid direct comparisons between different music cultures altogether, especially those involved in establishing cultural evolution (Nettl, 2010, p. 70–92). This revisionism relies on a systemic over-evaluation of the scope and limitations of purely emic approach to the study of music, in combination with a drastic under-evaluation of the advantages of the etic approach (see Nikolsky et al.).

The special issue of *"The World of Music,"* the journal of the International Music Council, (Vol. 22, No. 3, 1980), was dedicated to the discussion of the possibility to draw a general history of music, explore its origins, and formulate generalities in evolution of different musical cultures. Its overwhelming negative conclusions reflected the emerging trend of replacing the comparative studies that had been conducted within the field of systematic musicology by the disconnected studies of each musical culture, presumably entitled to its own unique "history" (Myers, 1993). Accordingly, the official name of the discipline was changed from "systematic musicology" to "ethnomusicology." Noteworthy is the radical inversion of the views on comparative studies by the "old-timers," such as Bruno Nettl, from positive in 1968 (Nettl and Blum, 1968) to negative in 2005 (Nettl, 2005).

Responding to Gourlay's call for ethnomusicology to abandon the "pretense of objectivity" in favor of "humanization" (Gourlay, 1982), Western ethnomusicologists started viewing their mission as "the study of people making music" rather than "the study of music" as the term "musicology" suggests (Titon, 2015). They substituted the study of text (that has traditionally been considered the "primary reality" for studying the arts) with the study of people's behaviors, causing a methodologic shift away from the analysis of music to sociology (Zemtsovsky, 1997)—exemplified by the following quotation: "Unless the formal analysis begins as an analysis of the social situation that generates the music, it is meaningless" (Blacking, 1974, p. 71). Analysis *per se* has acquired the reputation of a form of "composing" the analyzed piece of music on the part of an analyst, thereby distorting the original meaning of that music (Agawu, 2004). In effect, such a position deprived Western ethnomusicology of an empirical and objective foundation.

This trajectory is polarly opposite to the trend of equipping musicology with means of scientific research, prevalent in the former Soviet Bloc countries (Myers, 1993)[3]. Integration

[2]This refers not only to the matters of music theory and the performance practices of Western forms of music, but to the undocumented implicit music theories transmitted orally by music-makers within non-Western traditional music cultures. Confused by scientific terminology, they also often see scientific approach as essentially foreign to their music practice—despite their appreciation of scholarly attention to their respective traditions.

[3]Perhaps, a good demonstration of the unity of scientific and humanitarian approaches to music, common in the Soviet and modern Russian academia, is the partnership between both authors of this article: A. Nikolsky started his career as a professional composer, music theorist, and pianist, whereas L. Perlovsky—as a nuclear physicist. Somehow both authors met in their research on the origins and evolution of music.

of scientific and humanitarian knowledge has been a long tradition there since Lobachevsky[4] (Grigoryan, 2011)—affecting scholarship of all orientations: materialistic (Aleksei Losev), neo-Christian (Pavel Florensky), and esoteric (Peter Ouspensky). Music-related studies made no exception: noteworthy were such figures as Losev[5], Maykapar[6], and Samoilov [7]. All major Russian researchers of psychoacoustics could interpret/perform music at a professional level or/and compose or arrange music: Sofia Beliayeva-Ekzempliarskaya, Nikolai Garbuzov, Boris Teplov, Aleksey Ogolevets, Alexander Volodin, Yevgenii Nazaikinsky, and Yuri Rags. In 1944, Institute of History of Arts at the USSR Academy of Sciences was founded for scientific investigation of the arts (including music).

In the West too, there are ongoing attempts to bridge humanities and science by making musicology "empirical" (Clifton, 1983; Deutsch, 1996; Gjerdingen, 1999; Clarke and Cook, 2004; Honing, 2006; Baily, 2009; Schneider and Ruschkowski, 2011; Kendall and Lipscomb, 2013).

After all, every musician routinely conducts informal experiments: performers evaluate alternative ways of rendering music; studio-musicians experiment with various arrangements, and ear-trainers test students. Experimental trial of the premises of musical theory was discussed at the year-long seminar at Stanford, involving representatives of humanitarian and scientific disciplines (e.g., Lerdahl, Narmour, Gjerdingen, Bharucha, Palmer, and Krumhansl)—with the outcomes published in a special issue (1996) of "Music Perception." Since then, this idea has attracted attention of many scholars..

This is the direction we pursue in this collection of papers. It starts with Harvey overviewing the origins of music and presenting a theory of music that reflects a "society of selves" through social cooperation. Harvey attributes the invention of music to the promotion of group coherence and personal well-being.

Montagu informs readers without music education about the capacities of early humans, their possible musical behavior, and the overall evolution of musicality. Special attention is given to the development of musical instruments, which provides a window to the reconstruction of the musicking practices of the past.

Malloch and Trevarthen present an account of music in terms of human cognition and biology, with emphasis on musical education. They show how cultural life and learning depend on the motivation for sharing projects of thought and action, musically. Music empowers the transmission of the narratives of one's "inner life" in bodily movements. This ongoing practice must have transformed the primate brain for the affective regulation of social learning, thereby determining the evolution of human musical mind.

Brown updates his widely acknowledged musilanguage theory (Brown, 2000) by proposing the joint prosodic origin of proto-language and proto-music, where both shared specialization in emotional communication and neither featured scaled pitches. Brown introduces a "prosodic scaffold" model—i.e., specific vocal articulations and accompanying mimics/gestures forming signs for "acoustic pantomimes," designed to express one's affective state. According to Brown, combinatorial and compositional mechanisms of utterances generated the affective prosody, designating characteristic patterns of global acoustic expression for common emotional states. This forged into national prosody that branched into proto-language and proto-music based on, respectively, dialogic and chorusing formats of communication—distinguished by different approach to timing. At this point, music acquired concise temporal organization and synchronicity of collective production/perception.

Nikolsky introduces the term "isophony" to refer to the tonal and rhythmic properties of the musilanguage system. This amendment, endorsed by Brown, corrects the mismatch between the psychoacoustic and musicological taxonomies of musical texture. Nikolsky formulates a set of clear structural distinctions between the most common textural types: heterophony, homophony, and polyphony—in comparison to "isophony."

van der Schyff and Schiavio overview the key positions in evolutionary musicology, identifying the problems of the nature-or-culture antithesis. They elaborate on the "biocultural" approach exemplified in the work by Tomlinson (2015), to resolve these problems. The authors examine a range of supporting evidence for this approach vs. the "embodied" approach that regards music as a bio-cultural process governed by interaction of the enacted and social aspects. The authors cross-relate the current developments in evolutionary musicology, such as "enactivism" and "4E cognition," and suggest how the biocultural and "enactivist" approaches can be improved.

Nikolsky brings together insights from semiotics, musicology, psychoacoustics, evolutionary biology, anthropology, ethology, linguistics, and geomusicology to coin a new line of inquiry revolving around the concept of expressive aspects of music—in contradistinction from phonetics and prosody in natural languages. Central in this approach is the role of tonal organization and the comparative analysis of the acoustic features

[4]Nikolai Lobachevsky outlined the principles of the ideal scholarship in his speech "On the most important subjects of education" (1828), which he delivered upon his election as the rector of the Kazan University.

[5]Losev's semiotic studies eventually became most influential in the Soviet academia—laying out an encyclopedic Aristotelean-like theoretic foundation for all the disciplines in the empirical sciences and the arts. Losev's books "Music as a Subject of Logic" (1927) and "Dialectics of Artistic Form" (1927) reflected his lectures in Moscow Tchaikovsky Conservatory, where he taught aesthetics, and synthesized the neo-Platonic concept of music as a sounding number with Romantic ideas of symbolism in coining rational definitions of the notions of rhythm, melody, and harmony.

[6]Maykapar (1900) was a renowned composer, an outstanding concert pianist, and a distinguished pedagogue. He pioneered Helmholtz' approach in Russia to validate and reform the traditional way of teaching and making music—which included the psychoacoustic investigation of attention, melodic intonation, rhythm, timbre, tonality, and modality (Maykapar, 1900).

[7]Alexander Samoilov, assistant of Ivan Pavlov, and one of the developers of electrocardiography, was a virtuoso pianist and a musicologist who authored a number of publications on music theory (i.e., "Natural numbers in music," "Musical ethnographic museum instruments and their musical tuning," "Musical notation and its history"). He founded the Scientific-musicological Circle at Moscow University in 1902, which was later transferred to the Moscow Conservatory, where Sergei Taneyev and Samuel Maykapar subsequently took over his directorship. After moving to Kazan, Samoilov taught courses in math, physics, acoustics, music theory, and music history, in Kazan University.

of indigenous musical traditions and animal communication calls. Nikolsky offers a new method of multifactorial modal analysis of tonal organization and its graphic representation ("musogram") and examines its pros and cons in light of the emic/etic antithesis. He argues that music evolved from animal communication through gradual substitution of "one-ended" communication with "two-ended," where each of multiple aspects of expression acquired a repertory of proprietary signs for effective communication of emotional information, thereby remapping the animal-like instinctive correspondences between acoustic traits and affective states. The complete "semiotization" of all principal aspects of expression must have occurred no earlier than during the Neolithic "revolution" within the framework of the emerging bi-specific communication between humans and domestic animals—exemplified in the surviving pastoral culture of kulning.

Jan proposes yet another method of musicological analysis for tracking the lineage of diachronic evolution of specific musical structures based on the memetic approach to cultural changes. Elaborating on the research by Savage (2016), Jan combines the quantitative corpus-analysis techniques, adapted from molecular biology, with qualitative method of identifying perceptual-cognitive elements of music—"musemes"—revealed by music's motivic organization. This novel humanitarian-scientific integration promises a compelling and potentially testable means for studying the cultural evolution of music.

Lumaca et al. define a new interdisciplinary field of research—the contribution of neural constraints and biases on the cultural evolution of musical structures through the chain of cultural transmissions. Capacities of the human brain constrain acquisition, production, and reproduction of music. To illustrate this, the authors demonstrate a progressive diatonization of tonal organization in the multi-generational signaling game settings, which suggests that the smaller the information-processing bottleneck in individuals, the larger the pressures to regularize the music material. This poses new intriguing questions, such as the *r*ole of neural variability in music diversity—which is of greatest value for folk forms of music that entirely rely on oral transmission.

Podlipniak explores the Baldwinian evolutionary modeling in search for a compelling and testable theory for the phenomenon of human musicality, hypothesizing that it might constitute an adaptive phenomenon. In the Baldwin effect, animals, learn new behaviors that allow them to survive and reproduce in the changing environment. Podlipniak argues that the increasing group size of ancestral hominins required new mechanisms of "social consolidation" in the form of "collective imitation." There is then a natural selection for the evolution of new circuitry for vocal learning to maintain social bonds, thereby reducing the cost of learning.

Nikolsky et al. describe and explain the "timbre-based music" as a special system of musicking, communication, psychological, and social usage, along with corresponding beliefs—placing it in the timeline of the evolution of music. Timbral music opposes conventional Western music by its personal orientation: musicking here occurs primarily for oneself and/or for close relatives/friends. Throughout northeastern Eurasia, "personal song" serves as an important means of individual identification and territorial marking, akin a passport, supporting individual's mental health under harsh environmental conditions. The authors use demographic, geomusicological, paleoenvironmental, organological, and paleophonological data to argue that Siberian timbre-oriented music is remnant of the pan-Eurasian prehistoric music tradition that originated from the Last Glacial Period. Western frequency-oriented music, with its reliance on collective production/perception, might have been exported from Africa (where the population was much denser) before the LGP.

Ravignani et al. draw a parallel between researching the evolutions of music and language. They call for a systemic revaluation of an empirical and scientific approaches as opposed to claims from Chomskyan and anthropological perspectives against the scientific study. Given the intellectual influence of Chomsky, this point is extremely important. It would benefit musicologists to follow the lead of phonologists in establishing the evolutionary chain of developments. Re-aligning musicological and linguistic methodologies can allow the developmental psychologists and ethologists to make better choices between musical and linguistic paradigms in their research.

Fenk-Oczlon reports an intriguing phenomenon: the number of vowels and pitch-classes in native languages and native music tend to match. This is most pronounced in cases of simple systems—tritonic musical vs. 3-vowel vocal systems—but also noticeable in very complex 12-element systems. The mean values also match at 5–7 elements. Such correspondence supports Brown's model by revealing the shared ground between musical and vowel pitches.

Ravignani and Madison analyze the phenomenon of isochrony across human music and speech against animal communication, integrating the data from mathematics, physics, signal processing, physiology, and neuroscience. They define the concept of isochrony and propose an evolutionary hypothesis to explain why amongst all animals it is only humans that possess superb isochronous perception which does not confer evolutionary advantage to modern humans.

Honing et al. use an EEG oddball paradigm to assess the neural sensitivity to isochronous or arrhythmic beats in two monkeys. This non-invasive EEG methodology enables a direct comparison of the perception of monkeys, non-human primates, and humans. The authors found the MMN responses to the isochronous pattern but no strong evidence for beat sensitivity, confirming the Gradual Audiomotor Evolution model (Merchant and Honing, 2014) that holds metric organization as a biological marker of human music.

Loui et al. raise an intriguing question of musical anhedonia through cross-examination of its rare case against a panel of neurotypical participants. Their findings demonstrate the categorically different decreased connectivity between auditory and reward systems, supporting the Mixed Origins of Music model (Altenmüller et al., 2013). The authors identify neural pathways engaged in music's operation as an affective signaling system.

Based on his extensive teaching experience and research, Crickmore asserts his earlier experimental study and the test settings for measuring "aesthetic emotions"—listeners' response to detected "musical emotions" expressed by music creators. His revised test paves the road for clarifying the relations between human emotions, genres, and personality.

Masataka demonstrates that young people with autism spectrum disorder display more interest to dissonant music than typically developed matched group. This indicates that neural diversity within autism spectrum might have played a role in the evolution of dissonant sad music—important for overcoming negative aesthetic emotions of cognitive dissonances (Masataka and Perlovsky, 2013).

Trulla et al. explain consonance of musical intervals based on "second order beats,"[8] described by an approach borrowed from dynamical systems analysis—a quantitative index obtained from Recurrence Quantification Analysis. The novelty of this method is that it accounts for frequency ratio relationships plus temporal behavior. The authors confirm that musical consonance/dissonance has a mathematical foundation and that music perception, in general, and harmonic intervals, in particular, are a consequence of the entrainment of the nervous system with sound excitation.

[8]Second order beats are perceptible when two pure tones slightly deviate from simple, small-integer ratio relationships, e.g., 2:1 (octave), 3:2 (perfect 5th), and 4:3 (perfect 4th). Deviation from simple mathematical relationships results in some degree of chaos, perceived in dissonant intervals, e.g., tritones, minor 7ths.

Keller et al. provide a piece of evidence to support the Darwinian hypothesis of music's origin in sexual selection. In their study, a professional boys' choir was found to exhibit musical behavior essentially similar to male chorusing in many animal species. In female presence, bass singers instinctively emphasized and rose their singing formant—despite the conventional non-acceptance of such technique in choir-singing. This alteration, however, resulted in a more expressive performance. The authors explain this by covert competition between sexually mature males for female attention, which inadvertently maximizes choir's collective output.

We hope that this volume will bring us closer to answer Darwin's quest (Darwin, 1890) to disclose the mystery of music's origin.

## AUTHOR CONTRIBUTIONS

AN wrote the first draft of the manuscript. LP provided critical revision of the manuscript and important intellectual contributions. All authors contributed to the article and approved the submitted version.

## ACKNOWLEDGMENTS

The authors would like to thank all the contributors of this Research Topic for their work and the high quality of the submitted articles.

## REFERENCES

Agawu, K. (2004). How we got out of analysis, and how to get back in again. *Music Anal.* 23, 267–286. doi: 10.1111/j.0262-5245.2004.00204.x

Altenmüller, E., Kopiez, R., and Grewe, O. (2013). "Strong emotions in music: are they an evolutionary adaptation?," in *Sound - Perception - Performance*, ed R. Bader (Heidelberg: Springer), 131–156. doi: 10.1007/978-3-319-00107-4_5

Babbitt, M. (2003). "The composer as specialist [Who Cares if You Listen?]," in *The Collected Essays of Milton Babbitt*, ed S. Peles (Princeton, NJ: Princeton University Press), 48–54. doi: 10.1515/9781400841226.48

Baily, J. (2009). Crossing the boundary: from experimental psychology to ethnomusicology. *Empir. Musicol. Rev.* 4, 82–88. doi: 10.18061/1811/37484

Becker, J. (1986). Is western art music superior? *Music. Q.* 72, 341–359. doi: 10.1093/mq/LXXII.3.341

Blacking, J. (1974). *How Musical Is Man?* Seattle, WA: University of Washington Press.

Bohlman, P. V. (1999). "Ontologies of music," in *Rethinking Music*, eds N. Cook and M. Everist (Oxford: Oxford University Press), 17–34.

Brown, S. (2000). "The 'musilanguage' model of language evolution," in *The Origins of Music*, eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 271–300. doi: 10.1037/e533412004-001

Chomsky, N. (2016). *What Kind of Creatures Are We?* New York, NY: Columbia University Press. doi: 10.7312/chom17596

Clarke, E. F., and Cook, N. (2004). *Empirical Musicology: Aims, Methods, Prospects*. Oxford, NY: Oxford University Press.

Clifton, T. (1983). *Music as Heard: A Study in Applied Phenomenology*. Cambridge, MA: Yale University Press.

Cohen, B. R. (2001). Science and humanities: across two cultures and into science studies. *Endeavour* 25, 8–12. doi: 10.1016/S0160-9327(00)01335-1

Darwin, C. (1890). *The Descent of Man and Selection in Relation to Sex, 2nd Edn*. London: John Murray.

Deutsch, D. (1996). Preface to the special issue. *Music Percept.* 13:263. doi: 10.2307/40286172

Dubal, D. (1985). *The World of the Concert Pianist*. London: Gollancz.

Gjerdingen, R. O. (1999). "An experimental music theory," in *Rethinking Music*, eds N. Cook and M. Everist (Oxford: Oxford University Press), 161–170.

Gould, S. J. (2003). *The Hedgehog, the Fox, and the Magister's Pox: Mending the Gap Between Science and the Humanities*. New York, NY: Harmony Books.

Gourlay, K. A. (1982). Towards a Humanizing Ethnomusicology. *Ethnomusicology* 26, 411–420.

Grigoryan, N. (2011). Science and art: their reciprocal advance (Наука и искусство: встречное движение). *Herald Russia. Acad. Sci.* 81, 463–467.

Honing, H. (2006). On the Growing role of observation, formalization and experimental method in musicology. *Empir. Musicol. Rev.* 1, 2–6. doi: 10.18061/1811/21901

Juslin, P., Friberg, A., Schoonderwaldt, E., and Karlsson, J. (2012). "Feedback learning of musical expressivity," in *Musical Excellence: Strategies and Techniques to Enhance Performance*, ed A. Williamon (Oxford: Oxford University Press), 247–270. doi: 10.1093/acprof:oso/9780198525356.001.0001

Kendall, R. A., and Lipscomb, S. D. (2013). "Experimental semiotics applied to visual, sound, and musical structures," in *The Psychology of Music in Multimedia* (Oxford, NY: Oxford University Press), 48–65. doi: 10.1093/acprof:oso/9780199608157.003.0003

Lerdahl, F. (1992). Cognitive constraints on compositional systems. *Contemp. Music Rev.* 6, 97–121. doi: 10.1080/07494469200640161

Masataka, N., and Perlovsky, L. (2013). Cognitive interference can be mitigated by consonant music and facilitated by dissonant music. *Sci. Rep.* 3:2028. doi: 10.1038/srep02028

Maykapar, S. (1900). *Musical Hearing, Its Significance, Specialty and Method of Correct Development* [Музыкальный слух. Его значение, природа, особенности и метод правильного развития]. Moscow: Jurgenson [Юргенсон].

McGinn, C. (2015). *Inborn Knowledge: The Mystery Within*. Cambridge, MA: MIT Press.

Merchant, H., and Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* 7:274. doi: 10.3389/fnins.2013.00274

Messner, G. F. (1993). Ethnomusicological research, another "performance" in the international year of indigenous peoples? *World Music* 35, 81–95.

Myers, H. (1993). *Ethnomusicology: Historical and Regional Studies*. New York, NY: Norton.

Nattiez, J.-J. (2012). Is the search for universals incompatible with the study of cultural specificity? *Hum. Soc. Stud.* 1, 67–94. doi: 10.2478/v10317-012-0005-2

Nettl, B. (2005). *The Study of Ethnomusicology: Thirty-One Issues and Concepts*. Champaign, IL: University of Illinois Press.

Nettl, B. (2010). *Nettl's Elephant: On the History of Ethnomusicology*. Champaign, IL: University of Illinois Press.

Nettl, B., and Blum, S. (1968). Towards the comparative study of the structure of traditional repertoires. *J. Int. Folk Music Council.* 20, 47–50.

Parncutt, R. (2017). The two cultures in German translation: humanities, science, and Wissenschaft. *Scott. Music Rev.* 4. Available online at: https://www.n-ism.org/SMR/Parncutt:_Wissenschaft.pdf

Regelski, T. A. (2014). Scientism in experimental music research. *Philos. Music Educ. Rev.* 4, 3–19. doi: 10.2307/40327103

Savage, P. E. (2016). *Measuring the cultural evolution of music: with case studies of British-American and Japanese folk, art, and popular music (dissertation)*. Department of Musicology, Tokyo University of the Arts, 120.

Savage, P. E., and Brown, S. (2013). Toward a new comparative musicology. *Anal. Approach. World Music* 2, 148–197. doi: 10.31234/osf.io/q3egp

Schneider, A., and Ruschkowski, A. von (2011). *Systematic Musicology: Empirical and Theoretical Studies*. Frankfurt am Main: Peter Lang Verlag. doi: 10.3726/978-3-653-01290-3

Snow, C. P. (1964). *The Two Cultures*. Cambridge: Cambridge University Press.

Thomson, W. E. (2010). Empirical musicology review: serialist claims versus sonic reality. *Empir. Musicol. Rev.* 5, 36–50. doi: 10.18061/1811/46748

Titon, J. T. (2015). Ethnomusicology as the study of people making music. *Muzikoloski Zbornik* 51, 175–185. doi: 10.4312/mz.51.2.175-185

Tomlinson, G. (2015). *A million Years of Music: The Emergence of Human Modernity*. New York, NY: Zone Books.

Wallin, N. L., Merker, B., and Brown, S. (1999). *The Origins of Music*. Cambridge, MA: MIT Press.

Wilson, E. O. (1998). *Consilience: The Unity of knowledge*. New York, NY: Knopf.

Woody, R. H. (2000). Learning expressivity in music performance: an exploratory study. *Res. Stud. Music Educ.* 14, 14–23. doi: 10.1177/1321103X0001400102

Woody, R. H. (2004). 5 Misconceptions about scientific research in music education. *Teach. Music* 11, 28–33.

Zemtsovsky, I. (1997). An attempt at a synthetic paradigm. *Ethnomusicology* 41, 185–205. doi: 10.2307/852602

Zemtsovsky, I. (2002). The apologia of text [Апология текста]. *Music. Acad.* [Музыкальная академия] 4, 100–110. Available online at: https://mus.academy/articles/apologiya-teksta

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Nikolsky and Perlovsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

OPINION
published: 03 May 2017
doi: 10.3389/fpsyg.2017.00651

# The Measurement of Aesthetic Emotion in Music

*Leon Crickmore* *

*Retired, London, UK*

**Keywords: enjoyment of music, growth in appreciation, evolution of music, factor analysis, aesthetic emotion, dynamic gestalt**

In an earlier Opinion Article, Perlovsky challenged cognitive scientists to measure aesthetic emotions (Perlovsky, 2014). The present author claims to have done exactly that almost 50 years ago (Crickmore, 1968, 1973). Details of supportive factor analyses (Thomson, 1951) may be viewed in the attached Supplementary Material.

## A SYNDROME TEST OF MUSIC APPRECIATION

My syndrome hypothesis serves both as a functional definition of music appreciation, and as an operational definition of a dynamic gestalt. The hypothesis states that when a musical composition has been assimilated aesthetically, it will leave the listener feeling interested, happier, more relaxed, with a desire to remain quiet, satisfied and without any particular mental pictures. To cite a typical example of such a response from a student:

> Record 5 (Whistlin' Rufus: Chris Barber's Jazz Band)
> L7[1], I+, M+, T−, V−, S+, P0
> "Mind completely at rest and relaxed. Very enjoyable piece of music."

The psychological experience indicated by this syndrome may differ in intensity according to the profundity of the music, or to the listener's emotional state at the time. Occasionally, the experience can even reach the sublimity which T. S. Eliot describes in the third of his *Four Quartets:*

> "Music heard so deeply
> That it is not heard at all, but you are the music
> While the music lasts."

Lest any readers should consider that such an intense experience would be unlikely to be within the grasp of an 18-year-old technical student, let me quote an instance from a student. After listening to a recording of the second movement of Tchaikovsky's Sixth Symphony—a piece which another student described as the "worst record played yet"—this student responded with the full syndrome, adding:

> "I wasn't listening to the music but it was *me* inside. I was living it–this was especially noticeable with the climaxes."

Various factor analyses of my experimental data were carried out. These showed the Syndrome Test to be independent of the Maudsley Personality Inventory, the Wing Standardized Test of Musical Intelligence and Raven's Progressive Matrices (A–E)—a non-verbal test related to general intelligence. Also, when a number of the musical items were repeated later, the increase in the scores of students proved to be statistically significant, indicating that the test can also be used to measure growth in appreciation. This feature raises an issue that may be interpreted as a demonstration of the statistical unreliability of the test, as has been suggested by Clive Gabriel in our joint

[1]Liking 7 or 6 was included in the earliest versions of the test but later discarded.

OPEN ACCESS

***Edited by:***
*Leonid Perlovsky, Harvard University, USA*

***Reviewed by:***
*Stephan Thomas Vitas, Independent Researcher, Silver Spring, USA*

***Correspondence:**
*Leon Crickmore*
*crickmore176@talktalk.net*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *06 January 2017*
***Accepted:*** *12 April 2017*
***Published:*** *03 May 2017*

***Citation:***
*Crickmore L (2017) The Measurement of Aesthetic Emotion in Music. Front. Psychol. 8:651. doi: 10.3389/fpsyg.2017.00651*

article (Gabriel and Crickmore, 1977). This article describes the first use of the test with a totally random population, rather than with a group of students who had chosen to participate in a course of musical appreciation. Alternatively, such a result could be cited as evidence of the inappropriateness of traditional psychometric methodology for measuring the formation of dynamic gestalten, since such a process involves the integration of sub-wholes at a moment of ripeness (Maritain, 1959, pp. 132–134)[2]. The most frequent pattern of responses which Gabriel found was the postulated profile (G) with the exception of an incorrect response on scale P (P+ instead of P−). He called this response "a relaxed G." Yet on another occasion one of my students after listening to the first movement of Beethoven's Fifth Symphony wrote:

> "It is surprising too that when one's mind 'wanders' into mental pictures the actual beauty of the music escapes one and only surface music is absorbed."

## AESTHETIC EMOTIONS

Perlovsky states that "emotions related to knowledge have been called aesthetic since the time of Kant." I would contend, however, that since Kant held knowledge to be exclusively "discursive," that is to say the opposite of receptive and contemplative (Pieper, 1959, p. 32), this dogmatic assertion has contributed to the formulation of a methodology in psychology which has constrained cognitive scientists from attempting to measure aesthetic emotions. This postulate led Kant (1790) to rank music in the "lowest place" amongst the arts, and Pinker (1997) to dismiss music as "mere auditory cheese." Cross (1999, 2013) and Bannon (2016), however, both cite evidence indicating that Darwin believed, but deemed it inappropriate to say so at the time, that music's origins preceded those of language, which later emerged as an offshoot.

The concept of aesthetic emotion in academic literature seems to have been first posited by the aesthetician Bell (1914) and incorporated into cognitive psychology by Payne (1961, 1965). Later it was developed philosophically by Langer (1957). Langer's argues that music is both "a symbolism without assigned connotation" and "our myth of the inner life," a view that is wholly compatible with my syndrome hypothesis. Nevertheless, I prefer the greater metaphysical precision of the older explanation offered by Thomas Aquinas who distinguished between conceptual knowledge (*per cognitionem*) and knowledge from experience (*per connaturalitatem*). The aesthetic enjoyment of music belongs to the latter category (Crickmore, 1966).

Probably the best psychological theory that matches such philosophical views lies in Gestalt psychology, in the concept of a "dynamic gestalt," as expounded by Perls, Hefferline and Goodman's in *Gestalt Therapy* (Perls et al., 1951, pp. 403–404). Their book maps a process of contact, conceived as a single whole but which can also be conveniently divided into a series of grounds and figures: Fore-contact; Contacting; Final contact—when a listener would be "all ears" (Crickmore, 1972)—and Post-contact.

## MUSICAL EMOTIONS

Cooke (1959) has described music as a "language of the emotions," and has demonstrated how certain tonal patterns have become conventionally associated with particular basic emotions. The idea that an aesthetic response to music can be defined as a match between the emotions of a composer and those of the listener is widely accepted. However, people do in fact feel different emotions when listening to the same music, while a single listener may feel different emotions when listening to the same music at different times. I suggest, therefore, that musical emotions which do not lead to physical action are best considered as "as if" emotions. Langer has noted (Langer, 1957, p. 238), that music reflects "only the morphology of feeling." Musical emotions are more a matter of stipulation rather than of logical necessity.

Despite all that, a number of cognitive psychologists have assembled evidence for the measurability of a number of basic musical emotions (Juslin, 2013) With a view to future research, I have therefore added a means for registering these in a revised form of my Syndrome Test. The Revised Syndrome Test is shown in **Table 1**. The test in its original form is displayed in the upper part of the Table, with the exclusion of the opportunity for listeners to record their "Liking" for the music on a scale of 1–7. This was omitted, when it was discovered how highly the L. scale correlated with the I. scale. A new lower part of the test now provides opportunities for listeners to record the musical emotions they experience.

It has subsequently been found that the empirically verifiable musical emotions in several different cultures tend to be conventionally associated with a particular musical genre—for example, sadness—funeral music; pride—a national anthem. During the evolution of each of these genres, it seems likely that music would initially have been introduced as part of a communal ritual, thereby heightening the intrinsic emotions of a particular social activity. Only later, would a few well-manicure examples of these genres have come to be listened to in contemplative silence. Recently, some evidence has been found for a link between musical genre and personality (Neuman et al., 2016).

This evolutionary consideration brings to mind two even more significant pieces of recent pioneering research, the first by Perlovsky (2008, 2010, 2014) and Nikolsky (2015, 2016). For Perlovsky music has played a far greater part in human evolution than has so far been generally recognized. He defines its main evolutionary function as the resolution of the cognitive dissonances brought about by the split in the vocalizations of proto-humans into two types, one evolving into language, and the other into music. He also postulates the existence of a "knowledge instinct" through which the brain is able to match top-down and bottom-up signals in a manner for which he has coined the term "dynamic logic." Furthermore, his dynamic logic is computable, and could well share something in common with the psychological idea of a dynamic gestalt.

Working within Perlovsky's theoretical framework, and employing a methodology for melodic analysis little known

[2]I would justify the inclusion in the factor analyses of derived scores, and rotation to discover the most helpful viewing point on the grounds that loadings are simply mathematical entities (entia rationis).

**TABLE 1 | Revised syndrome test.**

| + | 0 | – | Responses |
|---|---|---|---|
| I Interested | Indifferent | Bored | |
| M Happier | No change | Sadder | |
| T More tense | No change | More relaxed | |
| V Desire to talk | No change | Desire to remain quiet | |
| S Satisfied | No change | Confused | |
| P Mental pictures | Name them: | | |
| | | | None 0 |
| | | | Some + |

Any other comments:

Then please tick any of the emotions listed below which you associate with the music you have just heard, and describe any others you have felt but which are not listed:

| | Tick ✓ |
|---|---|
| Joy | |
| Sadness | |
| Love | |
| Anxiety | |
| Anger | |
| Pride | |
| Any other: | |

*This revision of the syndrome test is characterized by the addition of a possibility for each listener to tick emotions felt during the music from a list of six emotions for which researchers have already found validating evidence. In this form the test has not yet been used experimentally. Until experiment provides some initial evidence concerning which emotions listeners actually do tick, and how frequently, it is not possible to decide whether, and if so how, such fresh data might be incorporated into a Factor Analysis of scores on the scales of the syndrome, together with those from some appropriate personality test. If those who believe that music is in some way a language of the emotions are right, such an analysis should be feasible. If my hypothesis that music is not a language of the emotions is true, any emerging data is unlikely to be compatible with factor analysis.*

outside the world of Russian-speaking musicologists, Nikolsky has managed to define, together with appropriate audio illustrations, 11 stages in the evolutionary development of music, from the time of Stephen Mithen's "Singing Neanderthals" (Mithen, 2005) and the earliest pre-modal music, via pentatonic and heptatonic stages and culminating in modern tonality. Could it be that, just as Pierre Teilhard de Chardin proposed (de Chardin, 1959; Crickmore, 1978), unlike the majority of modern scientists who tend to consider the biosphere as being the result of a largely fortuitous series of events, human evolution has a direction: namely the establishment of a phylum of "direct cerebralization?" Once this threshold of refection has been crossed evolution would continue in a new form, that of humanly invented combinations—logic, mathematics, art, and music. Could it also be possible that Nikolsky's model of tonal evolution might be taken to suggest a direction for musical evolution: toward the establishment of the diatonic system? Gary Tomlinson's *A Million Years of Music* (Tomlinson, 2015) is probably the best available survey of current knowledge in this field.

Each age in the evolution of music would have required a different style of listening. For us, in the twenty-first century, any description of this must take account of the effects of popular music with all its associated digitization and consumerist technologies. Our traditional definition of music in the West, based as it is largely on the world of concert halls and opera-houses therefore needs to be extended, possibly even to include listening to a Walkman and Muzak. Philip Tagg's book *Music's Meanings* (Tagg, 2013) could help us clear away some of the intellectual brushwood which has been accumulated over the years by musicologists and historians of music. Tagg has suggested an operational re-definition of music as "that form of human communication in which humanly organized non-verbal sound can, following culturally specific conventions, carry meaning relating to emotional, gestural, tactile, kinetic, spatial and prosodic patterns of cognition."

## CONCLUSION

There is a need for a programme of further research using my Revised Syndrome Test in **Table 1**. Such a programme should seek to clarify further the nature of the relationships between aesthetic emotions as measured by my test, musical emotions, genres, and personality. It should also include some fMRI scanning of the participants, both as they enjoy familiar music and as they grow to enjoy that which is less familiar to them. As long as Homo sapiens survive, music will survive too. In his poem *The Metalogue to the Magic Flute,* W. H. Auden has highlighted a matter which future researchers will have to bear in mind:

"The history of music as of Man
Will not go cancrizans, and no ear can
Recall what, when Archduke *Francis* reigned,
Was heard by ears whose treasure-hoard contained
A *Flute* already but as yet no *Ring,*
Each age has its own mode of listening."

## ETHICS STATEMENT

The experiment mentioned took place in the 1960s before the Helsinki agreement with the consent of the students involved.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2017.00651/full#supplementary-material

## REFERENCES


Bannon, N. (2016). Darwin, music and evolution: new insights from family correspondence on the descent of man. *Music. Sci.* 229–259. doi: 10.1177/1029864916631794

Bell, C. (1914). *Art*. London: Chatto & Windus.

Cooke, D. (1959). *The Language of Music*. Oxford: Oxford University Press.

Crickmore, L. (1966). Neo-thomism as a basis for the teaching of music. *Br. J. Educ. Stud.* 14, 36–44. doi: 10.1080/00071005.1966.9973165

Crickmore, L. (1968). An approach to the measurement of music appreciation, I & II. *J. Res. Music Educ.* 16, 291–301. doi: 10.2307/3344068

Crickmore, L. (1972). The musical gestalt. *Music Rev.* 33, 285–291.

Crickmore, L. (1973). A syndrome hypothesis of music appreciation. *Psychol. Music* 1, 21–25. doi: 10.1177/030573567312003

Crickmore, L. (1978). The complexity of music. *Teilhard Rev.* 13, 68–72.

Cross, I. (1999). "Is music the most important thing we ever did? Music development and evolution," in *Music, Mind and Science*, ed S. W. Yi (Seoul: Seoul National University Press), 10–39.

Cross, I. (2013). "Culture and evolution" in *Language, Music and the Brain,* ed M. A. Arbib, Strungmann Forum Reports Series, Vol. 10, ed J. Lupp (Cambridge, MA: MIT Press), 556.

de Chardin, P. T. (1959). *The Phenomenon of Man*. Transl. B. Wall. New York, NY; London: Wm. Collins Sons & Co Ltd.; Harper & Brothers.

Gabriel, C., and Crickmore, L. (1977). Emotion and music. *Psychol. Music* 5, 28–31. doi: 10.1177/030573567751004

Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Front. Psychol.* 4:596. doi: 10.3389/fpsyg.2013.00596

Kant, I. (1790). *Critique of Judgment*. Transl. by J. H. Bernard. London: Macmillan & Co.

Langer, S. (1957). *Philosophy in a New Key*. Cambridge, MA: Harvard University Press.

Maritain, J. (1959). *The Degrees of Knowledge*. London: Geoffrey Bles.

Mithen, S. (2005). *The Singing Neanderthals: The Origin of Music, Language, Mind and Body*. London: Weidenfeld and Nicolson.

Neuman, Y., Perlovsky, L., Cohen, Y., and Livshits, D. (2016). The personality of music genres. *Psychol. Music* 44, 1044–1057. doi: 10.1177/0305735615608526

Nikolsky, A. (2015). Evolution of tonal organisation in music mirrors symbolic representation of perceptual reality. Part I: prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.01405

Nikolsky, A. (2016). Evolution of tonal organisation in music optimizes neural mechanisms in symbolic encoding of perceptual reality. Part 2: ancient to seventeenth century. *Front. Psychol.* 7:211. doi: 10.3389/fpsyg.2016.00211

Payne, E. (1961). Emotion in music and in music appreciation. *Music Rev.* 22, 39–50.

Payne, E. (1965). Music, taste and personality. *Music Rev.* 26, 129–145.

Perlovsky, L. (2008). Music and consciousness. *Leonardo* 41, 420–421. doi: 10.1162/leon.2008.41.4.420

Perlovsky, L. (2010). Musical emotions: functions, origins, evolution. *Phys. Life* 7, 2–27. doi: 10.1016/j.plrev.2009.11.001

Perlovsky, L. (2014). Mystery in experimental psychology, how to measure aesthetic emotions? *Front. Psychol.* 5:1006. doi: 10.3389/fpsyg.2014.01006

Perls, F. S., Hefferline, R. F., and Goodman, P. (1951). *Gestalt Therapy*. New York, NY; London: The Julian Press Inc.; Sir Isaac Pitman & Sons Ltd.

Pieper, J. (1959). *Leisure as a Basis of Culture.* London: Faber & Faber.

Pinker, S. (1997). *How the Mind Works*. New York, NY: Norton.

Tagg, P. (2013). *Music's Meanings.* New York, NY; Huddersfield: The Mass Media Music Scholars Press.

Thomson, G. (1951). *The Factorial Analysis of Human Ability, 5th Edn.* London: University of London Press.

Tomlinson, G. (2015). *A Million Years of Music: The Emergence of Human Modernity*. New York, NY: Zone Books.


**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Crickmore. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Sociology

HYPOTHESIS AND THEORY
published: 20 June 2017
doi: 10.3389/fsoc.2017.00008

# How Music and Instruments Began: A Brief Overview of the Origin and Entire Development of Music, from Its Earliest Stages

*Jeremy Montagu**

*University of Oxford, Oxford, United Kingdom*

Music must first be defined and distinguished from speech, and from animal and bird cries. We discuss the stages of hominid anatomy that permit music to be perceived and created, with the likelihood of both *Homo neanderthalensis* and *Homo sapiens* both being capable. The earlier hominid ability to emit sounds of variable pitch with some meaning shows that music at its simplest level must have predated speech. The possibilities of anthropoid motor impulse suggest that rhythm may have preceded melody, though full control of rhythm may well not have come any earlier than the perception of music above. There are four evident purposes for music: dance, ritual, entertainment personal, and communal, and above all social cohesion, again on both personal and communal levels. We then proceed to how instruments began, with a brief survey of the surviving examples from the Mousterian period onward, including the possible Neanderthal evidence and the extent to which they showed "artistic" potential in other fields. We warn that our performance on replicas of surviving instruments may bear little or no resemblance to that of the original players. We continue with how later instruments, strings, and skin-drums began and developed into instruments we know in worldwide cultures today. The sound of music is then discussed, scales and intervals, and the lack of any consistency of consonant tonality around the world. This is followed by iconographic evidence of the instruments of later antiquity into the European Middle Ages, and finally, the history of public performance, again from the possibilities of early humanity into more modern times. This paper draws the ethnomusicological perspective on the entire development of music, instruments, and performance, from the times of *H. neanderthalensis* and *H. sapiens* into those of modern musical history, and it is written with the deliberate intention of informing readers who are without special education in music, and providing necessary information for inquiries into the origin of music by cognitive scientists.

**Keywords: music, rhythm, dance, instruments, development, social cohesion, performance, tonality**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky,*
*Braavo! Enterprises, United States*

***Reviewed by:***
*Dean Falk,*
*Florida State University, United States*
*Josef Verba,*
*The Muse of Life Society, Inc., United States*

***Correspondence:**
*Jeremy Montagu*
*jeremymontagu@gmail.com*

***Specialty section:***
*This article was submitted to Evolutionary Sociology and Biosociology, a section of the journal Frontiers in Sociology*

***Received:*** *01 March 2017*
***Accepted:*** *23 May 2017*
***Published:*** *20 June 2017*

***Citation:***
*Montagu J (2017) How Music and Instruments Began: A Brief Overview of the Origin and Entire Development of Music, from Its Earliest Stages. Front. Sociol. 2:8. doi: 10.3389/fsoc.2017.00008*

## HOW DID MUSIC BEGIN? WAS IT *VIA* VOCALIZATION OR WAS IT THROUGH MOTOR IMPULSE?

But even those elementary questions are a step too far, because first we have to ask "What is music?" and this is a question that is almost impossible to answer. Your idea of music may be very different from mine, and our next-door neighbor's will almost certainly be different again. Each of us can only answer for ourselves.

Mine is that it is "Sound that conveys emotion."

We can probably most of us agree that it is sound; yes, silence is a part of that sound, but can there be any music without sound of some sort? For me, that sound has to do something—it cannot just be random noises meaning nothing. There must be some purpose to it, so I use the phrase "that conveys emotion." What that emotion may be is largely irrelevant to the definition; there is an infinite range of possibilities. An obvious one is pleasure. But equally another could be fear or revulsion.

How do we distinguish that sound from speech, for speech can also convey emotion? It would seem that musical sound must have some sort of controlled variation of pitch, controlled because speech can also vary in pitch, especially when under overt emotion. So music should also have some element of rhythm, at least of pattern. But so has the recital of a sonnet, and this is why I said above that the question of "What is music?" is impossible to answer. Perhaps the answer is that each of us in our own way can say "Yes, this is music," and "No, that is speech."

Must the sound be organized? I have thought that it must be, and yet an unorganized series of sounds can create a sense of fear or of warning. Here, again, I must insert a personal explanation: I am what is called an ethno-organologist; my work is the study of musical instruments (organology) and worldwide (hence the ethno-, as in ethnomusicology, the study of music worldwide). So to take just one example of an instrument, the ratchet or rattle, a blade, usually of wood, striking against the teeth of a cogwheel as the blade rotates round the handle that holds the cogwheel. This instrument is used by crowds at sporting matches of all sorts; it is used by farmers to scare the birds from the crops; it was and still is used by the Roman Catholic church in Holy Week when the bells "go to Rome to be blessed" (they do not of course actually go but they are silenced for that week); it was scored by Beethoven to represent musketry in his so-called Battle Symphony, a work more formally called *Wellingtons Sieg oder die Schlacht bei Vittoria*, Op.91, that was written originally for Maelzel's giant musical box, the Panharmonicon. Beethoven also scored it out for live performance by orchestras and it is now often heard in our concert halls "with cannon and mortar effects" to attract people to popular concerts. And it was also, during the Second World War, used in Britain by Air-Raid Precaution wardens to warn of a gas attack, thus producing an emotion of fear. If it was scored by Beethoven, it must be regarded as a musical instrument, and there are many other noise-makers that, like it, which must be regarded as musical instruments.

And so, to return to our definition of music, organization may be regarded as desirable for musical sound, but that it cannot be deemed essential, and thus my definition remains "Sound that conveys emotion."

## SO NOW WE CAN ASK AGAIN, "HOW DID MUSIC BEGIN?"

But then another question arises: is music only ours? We can, I think, now agree that two elements of music are melody, i.e., variation of pitch, plus rhythmic impulse. But almost all animals can produce sounds that vary in pitch, and every animal has a heart beat. Can we regard bird song as music? It certainly conveys musical pleasure for us, it is copied musically (Beethoven again, in his *Pastoral Symphony*, no.6, op. 68, and in many works by other composers), and it conveys distinct signals for that bird and for other birds and, as a warning, for other animals also. Animal cries also convey signals, and both birds and animals have been observed moving apparently rhythmically. But here, we, as musicologists and ethnomusicologists alike, are generally agreed to ignore bird song, animal cries, and rhythmic movement as music even if, later, we may regard it as important when we are discussing origins below. We ignore these sounds, partly because they seem only to be signals, for example alarms etc, or "this is my territory," and partly, although they are frequently parts of a mating display, this does not seem to impinge on society as a whole, a feature that, as we shall see, can be of prime importance in human music. Perhaps, too, we should admit to a prejudice: that we are human and animals are not…

So now, we can turn to the questions of vocalization versus motor impulse: which came first, singing or percussive rhythms? At least we can have no doubt whatsoever that for melody, singing must long have preceded instrumental performance, but did physical movement have the accompaniment of hand- or body-clapping and perhaps its amplification with clappers of sticks or stones, and which of them came first?

Here, we turn first to the study of the potentials of the human body. There is a large literature on this, but it has recently been summarized by Iain Morley in his *The Prehistory of Music* (Morley, 2013). So far as vocalization is concerned, at what point in our evolution was the vocal tract able to control the production of a range of musical pitch? For although my initial definition of music did not include the question of pitch, nor of rhythm, once we begin to discuss and amplify our ideas of music, one or other of these, does seem to be an essential—a single sound with no variation of pitch nor with any variation in time can hardly be described as musical.

Studies based on fossil remains of the cranium and jaw formation of the early species of homo suggest that while *Homo ergaster* from between two million and a million and a half years ago could produce some variation of pitch but perhaps without much breath control, *Homo erectus* may have had greater ability, and *Homo heidelbergensis*, and certainly its later development from around a million years ago into the common ancestor of *Homo neanderthalensis* and *Homo sapiens*, could certainly "sing" as well as we can, though of course we can have no evidence of whether they could control such ability, whether they used it, and if so to what extent. So we can say that vocalization, while absent from the capability of our cousins the great apes and of the early forms of *Homo*, could be as old as at least a million years. It would seem that *Homo heidelbergensis* had the muscular abilities, but perhaps not the full mental capacities and that it was not until *H. sapiens* arrived that all the requirements for vocalization were in place, both exported and imported, and possibly not even in the earliest stages of the evolution of *H. sapiens*. It is here that there is controversy over the relative musical abilities of *H. neanderthalensis* and *H. sapiens*, to which in due course we shall return.

Much of this work also discusses the origins of speech as well as that of music. The two processes seem to have much the same

physiological requirements, the ability to produce the various consonants and vowels that enable speech, and the ability to control discrete musical pitches. But this capacity goes far beyond the ability to produce sounds.

All animals have the ability to produce sounds, and most of these sounds have meanings, at least to their ears. Surely, this is true also of the earliest hominims. If a mother emits sounds to soothe a baby, and if such sound inflects somewhat in pitch, however vaguely, is this song? An ethnomusicologist, those who study the music of exotic peoples, would probably say "yes," while trying to analyze and record the pitches concerned. A biologist would also regard mother–infant vocalizations as prototypical of music (Fitch, 2006). There are peoples (or have been before the ever-contaminating influence of the electronic profusion of musical reproduction) whose music has consisted only of two or three pitches, and those pitches not always consistent, and these have always been accepted as music by ethnomusicologists. So we have to admit that vocal music of some sort may have existed from the earliest traces of humanity, long before the proper anatomical and physiological developments enabled the use of both speech and what we might call "music proper," with control and appreciation of pitch.

In this context, it is clear also that "music" in this earliest form must surely have preceded speech. The ability to produce something melodic, a murmuration of sound, something between humming and crooning to a baby, must have long preceded the ability to form the consonants and vowels that are the essential constituents of speech. A meaning, yes: "Mama looks after you, darling," "Oy, look out!" and other non-verbal signals convey meaning, but they are not speech.

The possibilities of motor impulse are also complex. Here, again, we need to look at the animal kingdom. Both animals and birds have been observed making movements that, if they were humans, would certainly be described as dance, especially for courtship, but also, with the higher apes in groups. Accompaniment for the latter can include foot-slapping, making more sound than is necessary just for locomotion, and also body-slapping (Williams, 1967). Can we regard such sounds as music? If they were humans, yes without doubt. So how far back in the evolutionary tree can we suggest that motor impulse and its sonorous accompaniment might go? I have already postulated in my *Origins and Development of Musical Instruments* (Montagu, 2007, p. 1) that this could go back as far as the earliest flint tools, that striking two stones together as a rhythmic accompaniment to movement might have produced the first flakes that were used as tools, or alternatively that interaction between two or more flint-knappers may have led to rhythms and counter-rhythms, such as we still hear between smiths and mortar-and-pestle millers of grains and coffee beans. This, of course, was kite-flying rather than a wholly serious suggestion, but the possibilities remain. At what stage did a hominim realize that it could make more sound, or could alleviate painful palms, by striking two sticks or stones together, rather than by simple clapping? Again we turn to Morley and to the capability of the physiological and neurological expression of rhythm.

The physiological must be presumed from the above animal observations. The neurological would again, at its simplest, seem to be pre-human. There is plenty of evidence for gorillas drumming their chests and for chimpanzees to move rhythmically in groups. However, apes' capacity for keeping steady rhythm is very limited (Geissmann, 2000), suggesting that it constitutes a later evolutionary development in hominins. Perceptions of more detailed appreciation of rhythm, particularly of rhythmic variation, can only be hypothesized by studies of modern humans, especially of course of infantile behavior and perception.

From all this, it would seem that motor impulse, leading to rhythmic music and to dance could be at least as early as the simplest vocal inflection of sounds. Indeed, it could be earlier. We said above that animals have hearts, and certainly, all anthropoids have a heartbeat slow enough, and perceptible enough, to form some basis for rhythmic movement at a reasonable speed. Could this have been a basis for rhythmic movement such as we have just mentioned? This can only be a hypothesis, for there is no way to check it, but it does seem to me that almost all creatures seem to have an innate tendency to move together in the same rhythm when moving in groups, and this without any audible signal, so that some form of rhythmic movement may have preceded vocalization.

## BUT WHY DOES MUSIC DEVELOP FROM SUCH BEGINNINGS? WHAT IS THE PURPOSE OF MUSIC?

There are four obvious purposes: dance, personal or communal entertainment, communication, and ritual.

Dance we have already mentioned, though we can never know whether rhythmic motion led to the use of accompaniment, or whether the use of rhythm for any work led to people moving rhythmically in a way that became dance. It is well accepted in anthropology that when people are working, or moving together, their movements fall into a rhythm, that people may grunt and make other noises into that rhythm. The grunts may move into something that verges on or morphs into song; the other noises may be claps or beating pairs of objects together (concussive) or beating one object on another (percussive). Such objects can only be idiophonic, such as sticks, stones, and other solid objects that require no additional features to help them make a sound, in the classificatory system for instruments (Hornbostel and Sachs, 1914). This is simply because to create a drum with a skin (membranophones) is a complex process, because a skin will not produce sound unless it is under tension.

There is no doubt whatsoever that rhythmic sound without any melodic input must be regarded as music. It appears in many cultures, even if rarely, and we have Varèse's *Ionization* to take as an example from our modern orchestral repertoire.

Our second purpose was personal or communal entertainment. Communal entertainment, to some extent, overlaps with dance and with rhythmic work; personal entertainment overlaps for the mother and baby, mentioned above, with communication, as does the traveler using an instrument to indicate to people or villages that he passes that his purpose is peaceful and that he is not a robber intent on purloining their property, a well-known

practice anthropologically but one that we can have no way to measure its antiquity.

Our third purpose, communication by musical means is again widespread. We have the "bush telegraph" in Africa and other parts of the world with slit drums and other instruments, the alphorn in Switzerland and in other mountainous or marshy regions, the conch in Papua New Guinea, as random examples of the use of an instrument to pass messages. We have the whistling language of the Canary Islands (*silbo*) and many other parts of the world, and the high vocal calls of other peoples as examples of non-instrumental music for the same purpose.

Our fourth purpose, ritual, is a well-known trap in archeology and anthropology. Any object, any practice that cannot otherwise be explained, is assigned as "ritual." But there seems to be no form of religion, to use that word in its widest sense, that does not attract music to its practices. And here, we have another conflict, again that between music and speech. Schönberg's "invention" of *Sprechgesang*, an interface between speech and music, was nothing new. Many forms of ritual chants would be difficult to notate precisely in pitch; the words are spoken but they are inflected up and down quasi-melodically. Some bardic narrative is also an example of this, while often breaking intermittently into song. In both cases, the musical inflection renders the text less boring and helps the speaker with his or her memory of the text. It is undoubtedly speech, for the meaning of the words is the essential part, but there is also the element of pitch variation that would make an ethnomusicologist claim it to be music even while the practitioner would often vehemently deny any such claim, especially within the stricter forms of Islam, those in which music is forbidden.

Seemingly more important than these fairly obvious reasons for why music developed is one for why music began in the first place. This is something that Steven Mithen mentions again and again in his book, *The Singing Neanderthals* (Mithen, 2005): that music is not only cohesive on society but almost adhesive. Music leads to bonding, bonding between mother and child, bonding between groups who are working together or who are together for any other purpose. Work songs are a cohesive element in most pre-industrial societies, for they mean that everyone of the group moves together and thus increases the force of their work. Even today "Music while you Work" has a strong element of keeping workers happy when doing repetitive and otherwise boring work. Dancing or singing together before a hunt or warfare binds the participants into a cohesive group, and we all know how walking or marching in step helps to keep one going. It is even suggested that it was music, in causing such bonding, that created not only the family but society itself, bringing individuals together who might otherwise have led solitary lives, scattered at random over the landscape.

Thus, it may be that the whole purpose of music was cohesion, cohesion between parent and child, cohesion between father and mother, cohesion between one family and the next, and thus the creation of the whole organization of society.

Much of this above can only be theoretical—we know of much of its existence in our own time but we have no way of estimating its antiquity other than by the often-derided "evidence" of the anthropological records of isolated, pre-literate peoples. So let us now turn to the hard evidence of early musical practice, that of the surviving musical instruments.[1]

This can only be comparatively late in time, for it would seem to be obvious that sound makers of soft vegetal origin should have preceded those of harder materials that are more difficult to work, whereas it is only the hard materials that can survive through the millennia. Surely natural materials such as grasses, reeds, and wood preceded bone? That this is so is strongly supported by the advanced state of many early bone pipes—the makers clearly knew exactly what they were doing in making musical instruments, with years or generations of experiment behind them on the softer materials. For example, some end-blown and notch-blown flutes, the earliest undoubted ones that we have, from Geissenklösterle and Hohle Fels in Swabia, Germany, made from swan, vulture wing (radius) bones, and ivory in the earliest Aurignacian period (between 43,000 and 39,000 years BP), have their fingerholes recessed by thinning an area around the hole to ensure an airtight seal when the finger closes them. This can only be the result of long experience of flute making.

So how did musical instruments begin? First a warning: with archeological material, we have what has been found; we do not have what has not been found. A site can be found and excavated, but if another site has not been found, then it will not have been excavated. Thus, absence of material does not mean that it did not exist, only that it has not been found yet. Geography is relevant too. Archeology has been a much older science in Europe than elsewhere, so that most of our evidence is European, whereas in Africa, where all species of *Homo* seem to have originated, site archeology is in its infancy. Also, we have much evidence of bone pipes simply because a piece of bone with a number of holes along its length is fairly obviously a probable musical instrument, whereas how can we tell whether some bone tubes without fingerholes might have been held together as panpipes? Or whether a number of pieces of bone found together might or might not have been struck together as idiophones? We shall find one complex of these later on here which certainly were instruments. And what about bullroarers, those blades of bone, with a hole or a constriction at one end for a cord, which were whirled around the player's head to create a noise-like thunder or the bellowing of a bull, or if small and whirled faster sounded like the scream of a devil? We have many such bones, but how many were bullroarers, how many were used for some other purpose?

So how did pipes begin? Did someone hear the wind whistle over the top of a broken reed and then try to emulate that sound with his own breath? Did he or his successors eventually realize that a shorter piece of reed produced a higher pitch and a longer segment a lower one? Did he ever combine these into a group of tubes, either disjunctly, each played by a separate player, as among the Venda of South Africa and in Lithuania, or conjointly lashed together to form a panpipe for a single player? Did, over

[1] All the known archaeological instruments that we have, up to the end of the Neolithic period, are listed in tables by Morley (2013), and many are illustrated and described in his text.

the generations, someone find that these grouped pipes could be replaced with a single tube by boring holes in it, with each hole representing the length of one of that group? All this is speculation, of course, but something like it must have happened.

Or were instruments first made to imitate cries? The idea of the hunting lure, the device to imitate an animal's cry and so lure it within reach, is of unknown age. Or were they first made to imitate the animal in a ritual to call for the success of tomorrow's hunt? Some cries can be imitated by the mouth; others need a tool, a short piece of cane, bits of reed or grass or bone blown across the end like a key or a pen-top. Others are made from a piece of bark held between the tongue and the lip (I have heard a credit card used in this way!). The piece of cane or bone would only produce a single sound, but the bark, or in Romania a carp scale, can produce the most beautiful music as well as being used as a hunting call. The softer materials will not have survived and with the many small segments of bone that we have, there is no way to tell whether they might have been used in this way or whether they are merely the detritus from the dining table.

We have many whistles made from an animal phalange or toe bone, blown between a pair of protrusions at one end, across a sound hole near the center. Two of them come from the Mousterian period of the Middle Paleolithic, over 50,000 years ago, and there are many from the Aurignacian down to the Magdalenian and later; most, but not all, are reindeer phalanges. D'Errico has warned us, though, that the "sound hole" on many of these look as though they were made from a carnivore bite (D'Errico et al., 2003). It was in the Mousterian period that the Neanderthals co-existed with *Homo Sapiens*; the latter arrived in Europe between fifty and forty thousand years ago (though far earlier in the Near East), whereas Neanderthals had long been established in Europe, perhaps as long as 200,000 years before. Whether any that were blown by humans were used for signaling, or whether they were also used for music we cannot know, but whistles are certainly regarded as musical instruments.

More controversially in this Mousterian period, and certainly associated with other Neanderthal remains, is the young cave bear femur from the Divje Babe cave in Slovenia, dated to around 60,000 years Before the Present (BP). This has two holes in it and what might be three others at the broken-off ends, two on one side and one on the other. The fragment of bone is just over 10 cm long and while many people have claimed it as a flute, for it can certainly produce several pitches when reproductions of it are blown, many others have claimed that the holes are the result of other carnivores gnawing it, especially at the ends. As for the two complete holes, some writers have claimed that they are just the right size, shape, and spacing to have been produced by bears, for whose presence in the cave there is ample evidence, nor does there seem to be any trace of any possible human work on the bone. There is a very considerable literature on this possible instrument, well summed up and cited by Morley and by D'Errico et al., and the general consensus had been that it was not a musical instrument but simply the result of animal action. Nevertheless, the original discoverers have returned to the attack with a recent publication (Turk, 2014) which goes to show that human agency not only could have but did pierce those holes. For now, we can only leave this question open, with all the problems of an unicum; there are convincing conclusions on both sides of the argument, with at present rather greater weight on the "yes" side, partly due to this recent publication, and partly to the evidence in the following paragraph. What we really need are more examples from the Mousterian period.

This bone does raise the whole question of whether *H. neanderthalensis* knew of or practised music in any form. For rhythm, we can only say surely, as above—if earlier hominids could have, so could *H. neanderthalensis*. Could they have sung? A critical anatomical feature is the position of the larynx (Morley, 2013, 135ff); the lower the larynx in the throat the longer the vocal cords and thus the greater flexibility of pitch variation and of vowel sounds (to put it at its simplest). It would seem to have been that with *H. heidelbergensis* and its successors that the larynx was lower and thus that singing, as distinct from humming, could have been possible, but "seems to have been" is necessary because, as is so often, this is still the subject of controversy. However, it does seem fairly clear that *H. neanderthalensis* could indeed have sung. It follows, too, that while the Divje Babe "pipe" may or may not have been an instrument, others may yet be found that were instruments. There is evidence that the Neanderthals had at least artistic sensibilities, for there are bones with scratch marks on them that may have been some form of art, and certainly there is a number of small pierced objects, pieces of shell, animal teeth, and so forth, found in various excavations that can only have served as beads for a necklace or other ornamentation – or just possibly as rattles. There have also been found pieces of pigments of various colors, some of them showing wear marks and thus that they had been used to color something, and at least one that had been shaped into the form of a crayon, indicating that some reasonably delicate pigmentation had been desired. Burials have been found, with some small deposits of grave goods, though whether these reveal sensibilities or forms of ritual or belief, we cannot know (D'Errico et al., 2003, 19ff). There have also been found many bone awls, including some very delicate ones which, we may presume, had been used to pierce skins so that they could be sewn together. All this leads us to the conclusion that the Neanderthals had at least some artistic and other feelings, were capable of some musical practices, even if only vocal, and were clothed, rather than being the grunting, naked savages that have been assumed in the past.

It is in the Middle and Upper Paleolithic, from the Aurignacian period, which starts around 43,000 BP in eastern Europe and around 40,000 in the west, to the Magdalenian and later, ending around 10,000 BP, which we have a very considerable number of instruments, plus a few representations. Many of them, like those from Geissenklösterle above, are end-blown flutes made of bone, most commonly of large birds such as vultures and swans. Some of them are blown *via* a notch; some appear to be duct flutes, similar to our recorders, though of course the block made of wood, pith, or fiber has not survived—more probably, they are likely to have been tongue-duct flutes, using the tongue in the end instead of a block, and some are listed as such in Morley's tables—and others may have been plain-end blown, diagonally across the top, like the Arab *nay*. With these last, though, it is possible that a reed was used as the sound generator, either a double reed like that of our oboe or a split-cane single reed like that of

many Arab instruments, or possibly even lip-blown (trumpeted), though the narrowness of the bore makes this seem less likely. It is, therefore, probably better to refer to this last group as pipes, rather than as flutes.

Reproductions of many can be and have been played, but there is little to be learnt from this practice. We know what pitches and sounds *we* can get out of them, but unless we know their playing techniques, which of course we do not know, we cannot tell what sort of pitches and tone qualities they would have obtained in antiquity. Every recorder and tin-whistle player knows of a number of ways to inflect the pitch and the tone; every Arab *nay* player knows even more, and ethnomusicologists have produced evidence for even more, and our experimental musicians have shown that quite extraordinary pitches and sounds can be obtained from many of our orchestral instruments, sounds that their makers or normal players never conceived. Thus the archeologists (who are seldom trained musicians), who publish the scales and pitches of the pipes that they have found, can give us no more than conjecture and the experience of their own musicality. I have a collection of musical instruments from all over the world; I know the sounds that I can get out of them, but without the presence of the original player, or a field recording of the original player on that very instrument, I have no way to tell what sounds or pitches he or she produced. So much less can we have any idea what sounds and pitches were heard in the Paleolithic times.

However, there is one salient point, emphasized by D'Errico: a significant number of these pipes has varied spacing of finger holes. While, it would seem that the majority have the finger holes evenly spaced along the tube, there are certainly some that have a wider gap between the second and third holes. There are two fairly obvious possible reasons for this: one is that their "scale" of pitches had intervals similar to wholetones and minor thirds; the other that it was convenient or comfortable to have a wider gap between the two hands. This latter suggestion is raised because it was a standard feature of our flutes from the later Middle Ages right through into the early nineteenth century, and this was not only because from around 1700 the middle joint of the Baroque flute was divided into an upper and lower joint at this point – the earlier one-piece flutes also showed this gap. There are also some Aurignacian flutes or pipes that have one hole closer to another, showing that a semitone or a small wholetone was desired. Thus, these details emphasize that not only were these well-developed instruments, with the bodies well-scraped and smoothed, the finger holes with secure seating for the fingers, a certain amount of incised decoration, but that also there was a desire for precise tuning, and that they were not just made to produce fairly random pitches.

In addition, there is the point that many of these features appear both in Geissenklösterle in Germany, in Isturitz in France, in Spain, and also elsewhere, and over long periods of time, strongly suggesting that populations were not isolated but that there were links between them. This is not so surprising. If *H. sapiens* had traveled across Africa and into Europe, surely they could also travel between these areas and elsewhere.

There is little point in listing all these pipes; all the Paleolithic examples from Europe, or close by, found before 2013 are listed by Morley in his Appendices.

Were there other instruments? There is at least one conch trumpet, found in the Marsoulas cave, in the Haute-Garonne area of southern France, dating from around 20,000 years BP. Shell is a hard material that survives the ages, and although we have so far only this one example from the Upper Paleolithic, we have a very considerable number from the Neolithic times, some of them much further from the sea, so it is fair to assume a continuous use (Montagu, in press).[2] So what about animal horns? Here the material is soft, and only in very dry conditions such as desert sands do any survive; none of those that I have heard of or seen were blowing horns, but it seems likely that they existed. For blowing, the horn must be naturally hollow, such as those of the cow family, sheep and goats, antelopes, elephant tusks, hollow wood, gourds, and wide-bore bamboo, with the tip broken or cut off, or a hole bored in the side; such were surely blown in high antiquity (Montagu, 2014). There are several bullroarers from the Magdalenian period that we can be certain were instruments. There are many phalange whistles later than the Mousterian ones noted above. There are rasps, usually bones notched along their length, which would have been scraped with another bone or a stone for rhythmic music.

There is the complex of mammoth bones dating from around 20,000 BP, found in the Ukraine and published by Bibikov (1981). Many of the bones show signs of wear, almost certainly from repeated striking, and others, though this is not mentioned in the English summary, have striations similar to those of rasps, suggesting that some were scraped whereas others were struck. It is claimed that this was an ensemble, and although it would be difficult to prove that this was so, it would be even more difficult to show that each of these bones was struck only singly as an individual solo instrument. So here perhaps we have the first evidence of an "orchestra."

There are from the Magdalenian period, some 12,000 years BP, the caves themselves, where not only were stalactites struck but the caves themselves were used as resonators for sounds; both Lucie Rault and Lya Dams have brought together a number of convincing reports of this (Dams, 1985; Rault, 2000). Resonant stones must also have been struck outside the caves, the so-called rock gongs, boulders struck on resonant points, and these are of unknown antiquity but many bear well-worn cup marks on their surfaces. Rock gongs were first reported by Bernard Fagg in Nigeria, and following his article (Fagg, 1956), many more have been reported from around the world (Fagg, 1997).

There is no evidence in the Paleolithic period for stringed instruments nor for skin drums.

At what point in history did someone discover that by cupping the hands together and blowing between the knuckles of the thumbs produced a sound? This is a vessel flute or ocarina whose pitch is varied by moving the fingers to alter the area of open hole. Many peoples have long used gourds and other hollow vegetal objects, and today pottery, to play music in this way, also with the hands as hunting lures, but since there are no animal bones of such a shape, we can have no evidence of vessel flutes earlier than the Neolithic, in which period pottery first came into use.

[2]Montagu, J. (in press). *The Conch Horn*.

Did voice changers precede instruments? Did someone sing into a hollow object to change his voice from that of a human into that of a spirit or a deity? Was a shell sung into before ever a shell was blown? This precedence is something that has at times been suggested, but it can never be more than a hypothesis for we have no evidence to prove it. We do know that certain Greek statues had voice changers built in, usually a tube with a skin over one end, our kazoo, and there are many African masks with such a device.

Stringed instruments probably originated by the Mesolithic period, and certainly by the Neolithic, for it is in those periods that we begin to find flint arrow-heads, and the archer's bow and the musical bow are symbiotic as we shall see below (Balfour, 1899).

Skin drums (membranophones), as we said above, need the skin to be under tension to function. At what stage could there have been frames to which a skin could have been fastened securely enough to be tight enough to play? One can only say as early as skins were dressed, wetted, and dried on a frame, but since neither skins nor wooden frames, nor hollow logs, can ever have survived, this is simply an unknown; ceramic bodies rigid enough to support the skins can only have been available in or shortly before the Neolithic period.

So far, we have been discussing instruments only from Europe or its immediate environment. Simply, this is because where the evidence is. Archeology has been going on longer in Europe than elsewhere, as we have said. Much is being found now in China, but since most of it has been published in Chinese, much of this information is inaccessible, at least to me.

All the instruments that we have discussed above continued through the Neolithic and, with archery and pottery available, many others have joined them.

The earliest stringed instrument is undoubtedly the musical bow (Balfour, 1899). The one string instrument that might possibly be earlier is one that is identical with an animal trap—a noosed cord, presumably gut or sinew, running from a bent stick or branch to a peg in the ground. When an animal puts its head or leg into the noose, the cord is jerked from the peg and the stick or branch springs up and traps the animal. It has been suggested by Sachs, Balfour, and others that the hunter may also have plucked the string, so creating the ground bow, varying the tension of the cord, and thus the pitch, by bending the stick or branch. The ground harp is of unknown antiquity—our only evidence for the existence of the instrument is nineteenth-century reports from anthropologists.

Bows themselves, of course, never survive, but the presence of arrowheads in the lithic evidence proves their existence. Whether the archer's bow preceded the musician's or vice versa is arguable, but man's addiction to warfare, and even more to hunting, makes the archer's the more likely. We have ethnographic evidence for the use of the same bow for both purposes by the same person, but each developed in different ways, the archer's for strength and the musician's for producing musical sounds in different ways. The string of the musical bow is most commonly tapped by a light stick, initially presumably by an arrow, and is held to the player's mouth where, by changing the shape of the mouth, different overtones are sounded as with the jews harp (better and less prejudiciously called trump, which is the earlier English name). By dividing the string with a loop of cord linking the string to the stave, or by shortening the string at one end by the thumb of the holding hand, two fundamentals, each with their own overtones, makes a much greater range of pitches available. Attaching a gourd resonator to the stave creates greater volume, and opening or closing the mouth of the gourd against the player's chest will again elicit overtones. Both these forms survive to the present day in various modifications and many parts of the world, especially in Africa south of the Sahara (Kirby, 1934). A third form consists of attaching several bows to one resonator to form a pluriarc, as is still found in Central Africa.

One can postulate developments from both the gourd bow and the pluriarc. The gourd, eventually of wood, can be built on to one end of the stave to create both the category of instruments called lutes, with a straight stave as the neck, and of harps, with a curved stave. If the two outermost bows of the pluriarc become rigid, with a cross bar running between them to hold the distal ends of the strings of the inner bows, which then become redundant, the instrument is then much more stable and is called a lyre. Whether such developments took place, or whether lutes, harps, and lyres were independently invented, we can never know, but my own guess, based partly on various intermediate forms in various cultures, is for this process of development.

As for drums, frame drums are still ubiquitous around the world today, not only with our own tambourine, but a wooden or pottery body of manifold shapes exists almost everywhere. One possible early source for another type of drum is created by fixing the skin of the animal just eaten, over the top of the pot in which it had been cooked, so creating the instrument very appropriately called the kettledrum, using the word kettle in the sense of a caldron.

Another very common use of pottery is to create a rattle, a vessel containing seeds, pebbles, or nodules of pottery. Such vessel rattles must have been long preceded by gourds or woven leaves or baskets, all of which are still common today.

Once humanity entered the metal ages, the potentialities of instruments becomes infinite.

We can never know to what extent any groups of instruments or voices played together in high antiquity, though the existence of the group of mammoth bones above, does strongly suggest an ensemble. Not until the days of representational iconography, as in Mesopotamia and Egypt, or with the introduction of literacy, such as our Bible, do we have any real evidence. We have plenty of information from these sources.

What then did music sound like? We have early notations from Sumeria (Galpin, 1936) and Ancient Greece, the well-known hymn to Apollo, covering a wide range of pitches; Hickmann tried to derive a notation from hand-signals, called cheironomy, portrayed in Egyptian paintings and carvings (Hickmann, 1961). It has been thought by ethnomusicologists that less-advanced cultures than those, used pentatonic scales (five steps to the octave) such as we can still hear today in some areas, and perhaps even fewer steps with or without knowledge of the octave. But for these, naturally there is no evidence. Even with Sumerian, Greek,

and Egyptian systems, the various transcriptions of which are all controversial, we cannot know the actual sounds, for not until the later classical Greek period do we have written evidence of the sizes of scalar steps.

We do know, from the transcription of cuneiform tablets, that it was the Babylonians, and very possibly the Sumerians before them, who cataloged the skies and their constellations, establishing thus the basics of the calendar and of time that we use today, and who invented the hexadecimal system of mathematics. They turned their attention to sound also, and the Sumerians developed a system of diatonic scales based on alternating fourths and fifths. The Greeks, who took such knowledge from them, devised a diatonic scale based on the ratios of the harmonic series, starting from the eighth partial, a scale today called Just Temperament, one that is still used today by unaccompanied voices and sometimes by bowed string players or wind instruments playing without keyboards. For other instruments, such as lyres and harps, Just Temperament could also serve well, but only and until the players wished to change key; as soon as they did so, for reasons more complex than are needed here but are discussed below, chaos would ensue. Nevertheless, despite the purity of such a scale, we know that even the Greeks used other and more complex scales (Barbour, 1951) as, from the anthropological record, did many other peoples. Therefore, despite such transcriptions as we have of the ancient texts above, we can have no certain knowledge of what the music sounded like, for we do not know the exact sizes of the steps of the scales.

Even within Europe the 13th partial, the so-called alphorn fa, halfway between F and F-sharp appears in vocal music and on bagpipes as well as on natural horns and trumpets; the neutral third, between E and E-flat also appears, and as we shall see, the third is the most mutable interval in our classical music. In the Balkans, people sing in close seconds rather than wider intervals or unisons.

One thing that the ethnomusicologists can tell us is that either humanity has no inbuilt sense of consonant tonality, or that other people's sense of consonance is different from ours. The musical bow will by its nature produce the pitches of Just Temperament, for all its pitches are the overtones of the harmonic series, but despite this some peoples, who use the bow, will sing in seven equal steps to the octave. The one interval that does seem to be common to almost all peoples is the octave; this most probably originates with men and women singing in "unison" together, for women's voices tend to be an octave higher than men's. It is also a natural step to recognize when any piece of music extends beyond the range of one octave, and this repetition of scalar steps beyond the octave is built into many woodwind fingering systems.

We have many other examples of other scales that do not use what we, in our culture, may consider to be pure tuning. Let us take just one example that may be familiar to many of us today, the Javanese gamelan. This uses two different scales, *slendro* and *pelog*. Both employ the octave, but neither uses a pure fifth or third, the notes that make up our "common chord." *Slendro* has five almost equal steps to the octave; *pelog* has seven rather less equal steps. Not one of the steps of *slendro* is the same as those of *pelog*. Nor were the *slendro* or *pelog* in Java exactly the same between one gamelan and another, though similar, before the recent days when almost all gamelans are tuned to the pitches used by Radio Yogyakarta.

Nor are the scales of the Near and Middle East compatible with ours (Wizārat al-Tarbiyah wa-al-Taʿlīm, 1934). Nor even, save for the octave, are the pitches of Just Intonation the same as those of the Equal Temperament that we use on our pianos today. Each culture develops the tuning system that best suits its ideas of musicality. It is up to the cognitive scientists to determine why this should be so, but they have to admit, if they are willing to listen to the exotic musics of the world, that these differences exist.

Let us now return to the history of music and of the instruments on which it was played.

At least we do know what instruments some peoples used in the later millennia BCE, for not only do we have a few survivals in our museums from the Sumerian, Babylonian, Egyptian, Greek, Etruscan, and Roman periods, and also from the Orient, but we also have a wealth of iconography, much of it published in the *Musikgeschichte in Bildern* series by the Deutsche Verlag für Musik in Leipzig from the 1960s onward. This series is, alas, incomplete, for its publication ceased with the reunification of Germany.

We see among the Sumerians and Babylonians lyres and harps of various kinds, the latter quite small, a horizontal or vertical sound box with, at the distal end, a forepillar standing up at 90°, whereas in Egypt harps were normally curved, some of them as tall as the player, others, called the bow harp, were small enough to be held on the shoulder, and these last gradually passed into Central Africa where they are still found today. We see also lutes, a hollowed sound box like a small trough, with the open top covered with a skin to form the belly. A rod acts as the neck and passes through slits in the skin to hold it in place. These also still appear in Africa today. All these instruments were plucked, either with the fingers or a plectrum—the bow, such as we use on our fiddles, was as yet far in the future. There were pipes, usually double, held one in each hand, though sometimes, especially later in Egypt, lashed together so that the fingers of each hand could reach across both pipes. There were occasional drums, some very large, and many forms of rattles. We also see many of these instruments combined into what appear to be ensembles. This use of bands of instruments is confirmed in literature, for example in chapter 3 of the book of Daniel in our Bible where, when all the instruments play together, all those present bow down to the deity. Again in the Bible (II Samuel 6), a band of instruments escorts the Ark of the Covenant to David's city, with David dancing before them to the scorn of his queen.[3] Beware, however, of the huge choirs and groups of instruments in the two books of Chronicles; this is a late account, written long after any of the events it records, and smacks strongly of a child's playground exaggeration: "my brother is bigger and better than yours."

In ancient Greece, the lyre and the double pipe, the *aulos*, predominated. Lyres came in three forms. The simplest, the *chelys* or *lyra*, had a tortoise-shell body with two vertical curved

[3]For descriptions of all the instruments, see Montagu (2001).

wooden rods or horns, set in the shell with a third rod running horizontally as the cross bar. The strings were attached at one end to the bottom of the shell and at the other were twisted with *kollopes*, strips of skin, and wound round the horizontal bar. These *kollopes* set firmly enough on the bar to hold a tuning, but could be turned on the bar to retune. This type of lyre was taught to, and used for after-dinner symposia, by all educated people. It traveled up the Nile to the Meroitic people, probably in the Hellenistic period, and eventually throughout East Africa, where it is still used today, with the skin *kollopes* replaced with strips of cloth and the tortoise-shell with a gourd or wooden body as the resonator, and a skin belly. A more elegant form of Greek lyre, with longer curved arms, was called the *barbiton*. The professional musician's version, the *kithara*, was much more elaborate, with a wooden box-body and with what appears to be some form of semi-mechanized tuning devices. All three had gut strings that were normally plucked with a plectrum of wood, bone, or ivory, and all three are seen on many Greek vases and statues.

The *aulos* was a reed-pipe, shorter and somewhat stouter than the Sumerian and Egyptian; whether with a double reed like that of the oboe or a single reed like that of early folk clarinets as in the Near East today, is much argued, but Schlesinger's illustrations clearly show both types, though probably more often with the double reed (Schlesinger, 1939). The *aulos* passed on to Rome, where it was known as the *tibia*, to which quite elaborate tuning mechanisms were applied, with rings that could be turned to close off one hole and open another slightly differently placed, so as to play in a different key or mode. There was also a single pipe, the *monaulos*, and that is still found today, with a large double reed, all down the Silk Road, from Turkey, Kurdistan, and Armenia to China, Korea, and Japan. Whether it traveled east from Greece, or whether it originated in Central Asia like a number of other instruments and then traveled both east and west, is debatable.

That several instruments originated in Central Asia, probably somewhere between Persia and the Caspian Sea, is undoubted. The gong started there and was known in the Near East by St Paul (I Corinthians 13:1) as *chalkos ēchon* (Montagu, 2001, 123). The Chinese encyclopedias said that they got the gong from the West, which also suggests a Central Asian origin. The long trumpet seems to have started there also and it spread across the whole of Asia and to Greece, Etruria, and Rome, and in the Middle Ages through to North Africa as *alnafir* and, with the Moors, up into Spain as the *añafil*, and thence into the rest of Europe, and with the Hausa down into Ghana and Nigeria as the *kakaki*.

According to Al Farabi the Arab *'ud*, that became the lute in medieval Europe, also originated there, and so, around the eighth century CE, did the fiddle bow (Bachmann, 1969). Initially, this was a rough stick or reed scraping the string, but it was not long before it was modified with the strands of horsehair that we still use today.

This at last allowed stringed instruments to produce a sustained sound, something that could emulate the human voice, as all wind instruments had been able to do ever since their introduction.

In the early thirteenth century, and probably a little earlier, there came a revolution of the instruments we used in Europe. This seems to have been due to the often-interrupted symbiosis of Moorish, Jewish, and Christian cultures in Spain, and possibly also with some effect from returning Crusaders from the Holy Land. A flood of new instruments appeared, as can be seen in the many miniatures of the *Cántigas de Santa Maria*, a series of poems written by Alfonso X, called El Sabio, the wise.[4] We see there the Arab *'ud* which became our lute, the small bowed fiddle, the *rebab*, which became our rebec, the reed-blown pipe the *zamr*, which became our shawm, the ancestor of our oboe, several types of bagpipe, harps with a forepillar, various zithers such as the *qanun* that became our canon and then the psaltery, the transverse flute, other types of lute that became our gitterns and eventually citterns and guitars, *alnafir* that became the Spanish *añafil* and our long trumpet, pipe and tabor, the pipe played with one hand and the tabor struck with the other, which became a standard one-man band from the Middle Ages into the sixteenth century, the timbre, a frame drum that became our tambourine, and the *naqqere*, two small kettledrums, our nakers, that hung low from the belt in front of the player, and eventually became our timpani. Within the ensuing century, these spread all over western Europe and can be seen in a great many medieval manuscripts, church carvings, and other sources.

We know little of the extent that these played together. There are some group scenes in the *Cántigas*, but mostly, the miniatures show either one instrument or two of the same sort tuning or playing to each other. We do see large groups of instruments in manuscripts of the following centuries, but these are mostly portrayals of biblical scenes or of texts such as psalm 150 and may not represent anything that actually happened in the Middle Ages.

Then, in the fourteenth century, came another revolution, this time an industrial one (Gimpel, 1988). All over Europe, there had been windmills and watermills, primarily for grinding grain, but often also for minor industrial purposes. Now came the idea of siting watermills under the arches of bridges on major rivers, where the flow of water, restricted by the pillars of the bridge, thus produced far greater force. This powered mills for working metals and, for our purposes, of drawing brass and iron wire to standard quality and in much finer gages than had been available earlier except in softer, and more costly, metals such as silver and gold. The result was strings for harps, psalteries, and dulcimers and thence to keyboard instruments, first the clavichord, which was a keyed development of the monochord, and then the harpsichord. All, as can be seen in the manuscript of Arnault de Zwolle from around 1440, were established by that date (Le Cerf and Labande, 1932).

The use of keyboards led to a revision of musical pitch and tuning. Just Temperament had served well for unaccompanied voices and some solo instruments, but its inadequacies had now become more apparent. If one depends on the partials of the harmonic series, their ratios makes it obvious that the step from

[4]Escorial Library, Madrid, Ms. T I 1 (sometimes T. J. 1).

8 to 9 is greater than that of 9 to 10. To avoid using sharps and flats, let us take these pitches as C for 8, D for 9, and E for 10. And for clarity let us use the musicologist's interval-measuring system of cents, analogous to the general use of millimeters for linear measurement. The major tone of 8–9 is 204 cents; the minor tone of 9–10 is 182 cents, and together these make up the third, C to E, of 386 cents. Now if we want to play in C major, all is well, but if instead, we want to start a scale on D, we are in trouble, for where we need a major tone we have only a minor tone. Voices have no trouble with this for they simply shift the D and the E, but for any instrument with strings such as those of a lyre, a harp, or keyboards, the player has to stop and retune all his strings. The problem was already recognized by the ancient Greeks, and it was allegedly Pythagoras who solved the problem and who decided to make all the wholetones the same size, with 204 cents for each. However, adding those together produces a wildly sharp third of 408 cents from C to E, which when used in a common chord with C and G was so intolerable that in the Middle Ages it was regarded as a dissonance. Thus the Pythagorean Temperament was intolerable on the new keyboard instruments, and the music theorist Pietro Aron devised a new temperament in 1523. He returned to the natural third of 386 cents and, taking its mean or average of 193 cents for each whole tone, created the Quarter-comma Meantone Temperament. To the modern ear, accustomed to the Equal Temperament of our piano, with its wholetones of 200 cents and semitones of 100 cents, these differences may seem small, but if one listens to music played in other temperaments, it really does sound different—even today a 400-cent third still sounds quite badly out of tune. This whole subject is quite complex and Barbour, 1951, or the article on Temperaments in the *New Grove Dictionary of Music*, will give fuller details.[5] The basic problem is that the natural fifth of 702 cents is incompatible with the octave of 1200 cents; if one piles up a sequence of fifths, C to G, G to D, D to A and so on, the series will never return to C, only to a B-sharp 22 cents higher than C. Somehow those 22 cents, called a comma, have to be brought back into the octave, and this is done, with greater or lesser success, by using one of the various so-called irregular temperaments.

We have been neglecting vocal music. This has continued unchecked through the ages. When and how choral music, in our modern sense of song, evolved we do not know, but it had certainly appeared by biblical times and by that of the Greek dramatists. While we have mentioned some early suggested musical notations, music was normally taught by rote or simply by listening to others and joining in. What, if any, types of harmony were used, other than singing in octaves, we cannot know for we have no notation system, other than those early ones mentioned above for a basic melody, until we reach the early church chants. Here, we meet Gregorian and other church chants. These appear initially to have been purely monophonic, with everyone singing in unison. The earliest notation, called neumes, shows musical movement rather than precise pitches, and can only have served as a reminder of how music, already learned by rote, was to proceed. What pitch the music started on would depend on the preferred vocal range of the singers. Not until the thirteenth century do we start to see music written on a staff, then usually on only four lines rather than our present five-line stave, and with a symbol to tell us which line is C, similarly to our own alto or tenor clefs.

By the end of the twelfth century, we have composers such as Perotin writing organum, two or more parallel lines a fifth, fourth, or octave apart, with some slight freedom for each line to ornament a little. Organum probably derives from the organ itself, for while the first organs, which appeared in Alexandria in the second century BCE, were purely monophonic, though with the ability to play a chord, the larger church organs of the ninth or tenth centuries CE, used a system called *Blockwerk*. This meant that each key, when depressed, sounded a chord, a group of fourths or fifths and octaves. We have vivid descriptions of the tenth-century organ of Winchester Cathedral in Britain (Perrot, 1971), and we have surviving pipes from the organ of Bethlehem from the eleventh century of the Latin Kingdom of the Crusaders; the groups of lengths of these pipes show that this organ must also have used *Blockwerk* (Montagu, 2005).

What about secular music? Here, our earliest manuscripts seem to be from the thirteenth century with Adam de la Halle and his contemporaries writing motets for singers, and with anonymous, usually monophonic, dance music. Early polyphony, music in more than one part, was normally based on a cantus firmus, or tenor, often derived from a church chant, around which other, more elaborate parts, were woven. Polyphony of this sort seems to have been a purely European development; other cultures then, and in many cases still, prefer a single line or monophony, or if singing in groups or a single line with accompaniment, using heterophony, people all singing much, but by no means exactly, the same. Later motets might have three or four independent lines, sometimes each with their own text, woven together. These, in the early Renaissance, led to the madrigals and thence to our various styles of choral music today.

How do we define public performance, and how far back does it go? If one defines it as making music where other people can hear you, it must be as early as music ever existed. Any dance, whether Australian corroborees, war or hunting dances, people dancing on the village green, or any other similar occasions, must have involved music of some sort—how else could people keep their movement together? Here, we return to the use of rhythm, and surely to that of concussion or percussion of some sort, whether just body or hand clapping or that of instruments.

The shaman has always used music of some sort, often to help to throw him- or herself into the necessary trance. The bard has always been a valued member of society—and has always chanted and sung his lays, and always to self-accompaniment on an instrument. All these were "public" performances, either deliberately or at the very least where other people could hear them. At what stage was music deliberately performed to a public? Dance again, of course, and in religious ceremonies. The Christian church could be considered to be the first concert hall, with all free to enter and to hear the chant and, as time went

[5]There is also a comparatively simple explanation available on my website, jeremymontagu.co.uk, as a download: Montagu (1990).

on, listening to the deliberately composed music for the Mass. The medieval mystery plays were enacted in front of or within the church, and these always included music and were designed deliberately to draw in the public and to show them aspects of their religion.

When did people pay to hear music? Surely, this is part of our definition of public performance. Bards were certainly paid, domestic ones with board and lodging and presumably some cash, and itinerant ones certainly with cash or its portable equivalent, and shamans and medicine-men or -women always with cash or its equivalent, for that was the only way to be sure of a cure rather than a curse.

Formal concerts are said to have begun in Italy with the *Accademia*, meetings of intellectuals and musicians, in the fifteenth century, and private groups of musicians and musically interested people proliferated in many places, coming together to hear their own members playing and/or singing, for example the German *Collegia*. Aristocratic courts had their own orchestras, often merely for prestige, but sometimes, because the prince was himself a composer and musician. All these were private occasions, with admission confined to their members, their friends, and their guests.

Public concerts, with people paying for admission, began first in England perhaps as extensions of the Elizabethan theaters, where again people paid for admission, and which had often included musical performances along with the plays. England had no princely courts such as were common in continental Europe, and it was the first country to grow a middle class educated enough at the many grammar schools to appreciate musical culture and wealthy enough to pay for its pleasures. John Banister, himself a musician, was the first to invite the public to come, pay, and hear his concerts in 1673, and he was famously followed by Thomas Britton, "the small coal man," who opened a room above his shop to paying customers in 1678 and continued to provide weekly concerts for 36 years. Very shortly afterward, the first hall designed for musical performance was opened in London. It seems that in other countries such public performances did not take place until into the eighteenth century, and then in theaters and other improvised places, or out of doors. It was not until 1781 that the Leipzig Gewandhaus was built, the first public concert hall on the Continent.

A more elaborate form of music, the opera, began also as a court entertainment, but it rapidly became a public entertainment for which people paid for admission, probably because the costs of mounting an opera are far greater than chamber or orchestral concerts, and the first public opera house opened in Venice in 1637.

This is as far as we need to go for Europe, but what of the rest of the world? We have historical records and encyclopedias of music for the high cultures of China and India. We have, through archeology, surviving instruments such as the great assembly of Marquis Yi of Zeng in Suizhou, Hubei Province of China (Falkenhausen, 1993; So, 2000).[6] This was found in his tomb of around 433 BCE and elsewhere a Chinese set of Neolithic period bone flutes was found and published widely. Through the treasures of the great Depository of the Shōsōin in Nara (Shōsōin Office, 1967), we know how the instruments of the Chinese Tang court passed to Japan, and through the work of Laurence Picken and his successors how the music of that court changed in Japan (Picken et al., 1981 ff). All this tells us nothing further of how music began, but it does tell us that music progressed and developed, analogously with our own, in the high cultures of the world.

But, we have little knowledge of how, or even whether, music developed and changed in the rest of the world. We have glimpses, patchily, through the ages due to the iconographical records of some areas that we have mentioned above. We know much that goes on today, thanks to those ethnomusicologists who have been working around the world since the latter part of the nineteenth century, and we are dependent on their work for evidence of any possible sort simply because much of the music and the performances they recorded or described has vanished within our own lifetimes due to the globalized transmission of music. But even with that evidence, to what extent can we project any of it back in time? We could suggest that before the days of European exploration of the rest of the world, from the fifteenth century onward, peoples in sub-Saharan Africa were so isolated within their individual areas that their musics never changed from one generation to another. But that is a nineteenth-century attitude, of the time when Europeans refused to believe that sites such as Great Zimbabwe could ever have been built by African peoples, before the recognition of the great metal workers of West Africa and the high artistic levels of the Nok people or of Benin. I believe that any form of back-projection would be dangerous, whether in Africa or anywhere else in the world. I think that we simply have to say that we do not know and to admit that if *H. sapiens* could progress to such an extent as we know that it did in Europe and the Middle and Far East, so it could have done elsewhere.

We do have to say that much traditional music is dying out around the world, driven out by the perceived "superiority" of so-called "Western" music. Throughout the world now, there are symphony orchestras, even more widely there are all the manifestations of pop and other such musics. Yes, Bach, Beethoven, Mozart, Stravinsky, and others produced great works of music, but so did those of other cultures, and those musics are vanishing and their cultural contexts are dying out and treasures are being lost. And yet tradition manages to cling on, especially in the areas of pop music. West African versions of all the manifold varieties of popular musics do not sound the same as the New York versions. What we hear as "World Music," although heavily influenced by Western instruments and practices, still retains its local connotations and styles. The Soviet idea was that the individual solo performer from the eastern provinces should be replaced with groups on a concert platform with orchestras of alto, tenor, and bass versions of his or her instrument, still played their own musics in modified versions of their own styles. Music is and always has been created by people. It changes with time, and the ease of travel from the days of trains and steamships, and especially now globalization, has accelerated the rate of change

[6]This was published fairly briefly as So (2000), and in much greater detail as Falkenhausen (1993).

from the nineteenth century onward. But travel, even on foot and in log canoes, has been with us since the Paleolithic and so has inventiveness. Change in music and change in instruments will always be with us, but traditions, however changed, will always survive.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## ACKNOWLEDGMENTS

The author is grateful to the reviewers for their insightful comments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://journal.frontiersin.org/article/10.3389/fsoc.2017.00008/full#supplementary-material.

## REFERENCES

Bachmann, W. (1969). *The Origins of Bowing and the Development of Bowed Instruments up to the Thirteenth Century (Transl. Norma Deane)*. London: Oxford University Press.

Balfour, H. (1899). *The Natural History of the Musical Bow: A Chapter in the Developmental History of Stringed Instruments of Music*. Oxford: Clarendon Press.

Barbour, J. M. (1951). *Tunings and Temperaments: A Historical Survey*. East Lansing; New York: Michigan State University Press; Reprinted Da Capo Press, 1972.

Bibikov, S. N. (1981). *The Oldest Musical Complex Made of Mammoth Bones [in Russian with English and French Summaries]*. Ukrainian SSR: Institute of Archaeology, Academy of Science.

Dams, L. (1985). Palaeolithic idiophones: descriptions and comparisons. *Oxf. J. Archaeol.* 4, 31–46. doi:10.1111/j.1468-0092.1985.tb00229.x

D'Errico, F., Henshilwood, C., Lawson, G., Vanhaeren, M., Tillier, A. M., Soressi, M., et al. (2003). Archaeological evidence for the emergence of language, symbolism, and music? An alternative multidisciplinary perspective. *J. World Prehist.* 17, 1–70. doi:10.1023/A:1023980201043

Fagg, B. (1956). The discovery of multiple rock gongs in Nigeria. *Man* 56, 17–18. reprinted in African Music 13 (1956). doi:10.2307/2794038

Fagg, M. C. (1997). *Rock Music*. Oxford: Pitt Rivers Museum, Occasional Paper on Technology, 14.

Falkenhausen, L. (1993). *Suspended Music: Chime Bells in the Culture of Bronze Age China*. Berkeley: University of California Press.

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi:10.1016/j.cognition.2005.11.009

Galpin, F. W. (1936). *The Music of the Sumerians and Their Immediate Successors the Babylonians and Assyrians*. Cambridge: Cambridge University Press, 1955. reprinted Strasbourg: University Press, Librairie Heitz.

Geissmann, T. (2000). "Gibbon song and human music from an evolutionary perspective," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 103–123.

Gimpel, J. (1988). *The Medieval Machine: The Industrial Revolution of the Middle Ages*, 2nd Edn. Aldershot: Wildwood House.

Hickmann, H. (1961). *Musikgeschichte in Bildern II/1*. Ägypten: Deutscher Verlag für Musik.

Hornbostel, E., and Sachs, C. (1914). Systematik der Musikinstrumente: ein Versuch. *Z. Ethnol.* 4/5, 553–590. Available at: http://network.icom.museum/fileadmin/user_upload/minisites/cimcim/documents/H-S_20classification_20final_20version_20_282013_29_20without_20editorial_20markings-2.pdf

Kirby, P. (1934). *The Musical Instruments of the Native Races of South Africa*. London: Oxford University Press.

Le Cerf, G., and Labande, E. R. (eds) (1932). *Les Traités d'Henri-Arnault de Zwolle et de divers anonymes (MS. B. N. Latin 7295)*. Paris: Auguste Picard. repr. Bärenreiter.

Mithen, S. (2005). *The Singing Neanderthals*. Cambridge, MA: Weidenfeld & Nicholson.

Montagu, J. (1990). *Tuning and Tempering*. Available at: http://www.jeremymontagu.co.uk

Montagu, J. (2001). *Musical Instruments of the Bible*. Lanham, MD: Scarecrow Press.

Montagu, J. (2005). *The Oldest Organ in Christendom*. Available at: http://www.jeremymontagu.co.uk

Montagu, J. (2007). *Origins and Development of Musical Instruments*. Lanham, MD: Scarecrow Press.

Montagu, J. (2014). *Horns and Trumpets of the World*. Lanham, MD: Rowman & Littlefield.

Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology, & the Origins of Musicality*. Oxford: Oxford University Press.

Perrot, J. (1971). *The Organ from its Invention in the Hellenistic Period to the End of the Thirteenth Century (Transl. Norma Deane)*. London: Oxford University Press.

Picken, L., et al. (1981). *Music from the Tang Court*. Oxford and Cambridge: Oxford University Press.

Rault, L. (2000). *Musical Instruments. A Worldwide Survey of Traditional Music-Making*. London: Thames & Hudson.

Schlesinger, K. (1939). *The Greek Aulos*. London: Methuen.

Shōsōin Office. (ed.) (1967). *Musical Instruments in the Shōsōin*. Tokyo: Nihon Keizai Shimbun Sha.

So, J. (ed.) (2000). *Music in the Age of Confucius*. Washington, DC: Smithsonian Institution.

Turk, I. (ed.) (2014). *Divje Babe I. Paleolitsko najdišče mlajšega pleistocena v Sloveniji II. del: Arheologija (Upper Pleistocene Palaeolithic site in Slovenia Part II: Archaeology)*. Ljubliana: Opera Instituti archaeologici Sloveniae, 29.

Williams, L. (1967). *The Dancing Chimpanzee: A Study of the Origins of Primitive Music*. London: André Deutsch.

Wizārat al-Tarbiyah wa-al-Taʿlīm. (1934). *Recueil des travaux du Congrès de musique arabe*. Boulac: Imprimerie Nationale.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Montagu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

ORIGINAL RESEARCH
published: 14 September 2017
doi: 10.3389/fpsyg.2017.01559

# Simultaneous Cooperation and Competition in the Evolution of Musical Behavior: Sex-Related Modulations of the Singer's Formant in Human Chorusing

*Peter E. Keller*[1,2]*, *Rasmus König*[2] and *Giacomo Novembre*[3]

[1] MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, NSW, Australia, [2] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, [3] Department of Neuroscience, Physiology and Pharmacology, University College London, London, United Kingdom

Human interaction through music is a vital part of social life across cultures. Influential accounts of the evolutionary origins of music favor cooperative functions related to social cohesion or competitive functions linked to sexual selection. However, work on non-human "chorusing" displays, as produced by congregations of male insects and frogs to attract female mates, suggests that cooperative and competitive functions may coexist. In such chorusing, rhythmic coordination between signalers, which maximizes the salience of the collective broadcast, can arise through competitive mechanisms by which individual males jam rival signals. Here, we show that mixtures of cooperative and competitive behavior also occur in human music. Acoustic analyses of the renowned St. Thomas Choir revealed that, in the presence of female listeners, boys with the deepest voices enhance vocal brilliance and carrying power by boosting high spectral energy. This vocal enhancement may reflect sexually mature males competing for female attention in a covert manner that does not undermine collaborative musical goals. The evolutionary benefits of music may thus lie in its aptness as a medium for balancing sexually motivated behavior and group cohesion.

**Keywords: music, vocal expression, singer's formant, evolution, non-verbal communication**

OPEN ACCESS

***Edited by:***
*Leonid Perlovsky, Harvard University and Air Force Research Laboratory, United States*

***Reviewed by:***
*McNeel Gordon Jantzen, Western Washington University, United States*
*Jessica Phillips-Silver, Georgetown University Medical Center, United States*

***Correspondence:**
*Peter E. Keller*
*p.keller@westernsydney.edu.au*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *21 March 2017*
***Accepted:*** *28 August 2017*
***Published:*** *14 September 2017*

***Citation:***
*Keller PE, König R and Novembre G (2017) Simultaneous Cooperation and Competition in the Evolution of Musical Behavior: Sex-Related Modulations of the Singer's Formant in Human Chorusing. Front. Psychol. 8:1559. doi: 10.3389/fpsyg.2017.01559*

## INTRODUCTION

Music is a social communicative art form whose pervasiveness across human cultures suggests convergent evolutionary origins (McDermott and Hauser, 2005; Fitch, 2006; Merker et al., 2015). Although, these origins remain as mysterious today as they were in antiquity (McDermott, 2008; Morley, 2013), the comparative study of communicative signaling in human and non-human species has identified two broad classes of factors—competitive and cooperative—that could have driven the evolution of musical behavior (Brown, 2000; Mithen, 2005; Cross and Morley, 2009; Hagen and Hammerstein, 2009; Merker et al., 2015).

Competition for mates and territory in a variety of non-human animal species (e.g., songbirds and humpback whales) has selected for music-like signaling behavior that increases an individual's prospects for reproduction and survival by attracting sexual partners and repelling rivals

(Catchpole and Slater, 1995; Marler, 2000; Fitch, 2006; Smith et al., 2008). In the case of birdsong, playback experiments with European starlings, wrens, and flycatchers have shown that females are attracted to audio recordings of conspecific male vocalizations (Eriksson and Wallin, 1986; Mountjoy and Lemon, 1991; Johnson and Searcy, 1996). Related work with dunnocks has found that female responsiveness is greatest when the recording is of an alpha male, the song is complex, and the female is in her fertile period (Wiley et al., 1991). Furthermore, studies of song production in great reed warblers have demonstrated that males sing longer, more complicated songs when advertising for females than when already paired (Catchpole, 1983). Playback of such complex male songs elicits increased sexual display behavior in females (Catchpole et al., 1986), and males who sing more complex songs attract more females and produce more offspring (Catchpole, 1986). Song-like signaling may also play a role in attracting potential mates in some marine mammals. For example, the incidence of singing in male humpback whales is greater in the presence of unescorted mother-calf pairs (in which the adult female provides a mating opportunity) than for mother-calf pairs who are already escorted by another male (Smith et al., 2008). By analogy, it has been proposed—originally by Darwin (1871)—that the genesis of human music lies in courtship displays that evolved through sexual selection fuelled by competition between individuals (Miller, 2000).

Cooperative accounts provide an alternative by focusing on benefits to the individual gained through group membership. Group music making is a vital part of social life across the world's cultures, with individuals frequently coming together to communicate emotions and expressive intentions through rhythmically coordinated body movements and sounds (Merriam, 1964; Blacking, 1973; Nettl, 1983). In other social animals, the demands of cooperating with conspecifics has selected for collective displays where the inter-individual coordination of signals in space and time facilitates pair bonding [e.g., gibbon "duets" (Geissmann, 2000) and whistle matching in dolphins (Janik, 2000)] and group cohesion [e.g., call-and-response acoustic signaling in bats (Chaverri et al., 2012)]. Social accounts of the origins of human music hold that rhythmic coordination between individuals engaged in collaborative musical activities similarly facilitates cooperative behavior by promoting interpersonal bonding and feelings of mutual affiliation, trust, and commitment (Brown, 2000; Huron, 2001; Kirschner and Tomasello, 2010; Tarr et al., 2014; Weinstein et al., 2016). Practices involving interacting with others through music thus buttress human society and culture by exploiting the basic capacity for rhythmic interpersonal coordination to foster pro-social behavior (Lomax and Berkowitz, 1972; Mithen, 2005; Cross, 2012; Morley, 2013; Launay et al., 2016).

Accounts based on competition or cooperation alone, however, are not sufficient for dealing with the extensive palette of musical behaviors in the human repertoire, ranging from soothing lullabies, through festive folksongs and magnificent ceremonial and "art" music, to aggressive war dances and rap battles (Merriam, 1964; Nettl, 1983; Clayton, 2009; Clayton et al., 2012). An alternative is that human music serves both cooperative and competitive functions (Brown, 2000). On this view, music does not function solely in sexual or natural selection at the level of the individual or as a vehicle for group selection, but may rather play roles at both the level of the individual and the group.

The motivation for this unifying view comes from multiple angles. To begin with, it has been argued that sexual selection alone is not viable because music is rarely used as an obvious courtship display and is not a sexually dimorphic trait in humans (Brown, 2000). Natural selection is more generally problematic because the precise sense in which music serves to adapt individuals to their environment, and thereby enhance one's chance of survival, is obscure (Fitch, 2006). It has, however, been countered that pure cooperative accounts are questionable because they rely on group selection, which some evolutionary theorists consider to be a weaker explanatory principle than individual selection (Miller, 2000).

To circumvent these drawbacks, unifying accounts of human musical behavior postulating multilevel selection of both cooperative and cooperative functions have been advanced. One such account holds that coordinated displays in music and dance are beneficial in communicating information about coalition quality within a group to potential allies and competitors outside the group (Hagen and Bryant, 2003). Another unified account posits that music can function as a reward system that reinforces vital ritualistic group behaviors under some circumstances and as courtship or fitness displays in others (Brown, 2000). A strong version of the hypothesis that music functions both cooperatively and competitively, which we advance here, is that music is capable of doing so simultaneously, thus supporting different forms of communication in parallel at the level of the group and the individual. This dual-function hypothesis is motivated by work on rhythmically coordinated communal displays in non-human species (Merker, 2000).

Although, group music making can be viewed as an exalted human activity, it is only one amongst several stunning examples of rhythmic communal signaling in nature (Bailey, 1991; Merker, 2000; Ravignani et al., 2014; Merker et al., 2015). Bioluminescent fireflies flash together (Buck, 1988; Branham and Greenfield, 1996), fiddler crabs collectively wave their oversized claws (Murai and Backwell, 2006), fish drum their swim bladders in concert (Locascio and Mann, 2011), and cicadas (Hoppensteadt and Keller, 1976), crickets (Walker, 1969; Snedden and Greenfield, 1998), and frogs (Whitney and Krebs, 1975; Partridge and Krebs, 1978) produce choruses of sounds wherein coordination ranges from strict synchrony to relatively flexible alternation (Greenfield, 2005; Ravignani et al., 2014).

These non-human displays, which are typically produced by congregations of males, fulfill functions related to courtship and conflict avoidance (Greenfield, 1994). By coordinating their signals, group members maximize the amplitude of their collective broadcast and thereby increase its salience and geographic reach, resulting in a "beacon" effect (Buck and Buck, 1966) that enhances the ability to attract distant females, repel territorial adversaries, and confuse predators (Greenfield, 2005). While such coordinated behavior appears cooperative, research on communal sexual displays suggests that, in a number of species, it arises from the operation of competitive mechanisms

by which individual males jam—i.e., block or interfere with—rival signals (Whitney and Krebs, 1975; Greenfield and Roizen, 1993; Greenfield, 1994).

On this account, males who signal faster than their neighbors when competing for female attention gain an advantage by producing relatively early signals that effectively mask following signals (Otte and Loftus-Hills, 1979; Greenfield, 1994). Precise temporal coordination between males nevertheless ensues when leading calls trigger a call in neighbor after a fixed phase delay or a rebound interval associated with inhibitory resetting of a neural pacemaker (Buck, 1988; Greenfield and Roizen, 1993), with globally stable patterns of interaction occurring when variations in individual signaling rates cause switches in momentary leadership (Greenfield et al., 1997). Group coordination under such circumstances is thus an epiphenomenon resulting from competitive mechanisms selected by general female preferences for male signals that contain greater energy due to delivery at a fast rate or high intensity (Greenfield, 2005).

The current case study tested the hypothesis that human music is likewise capable of simultaneously serving cooperative and competitive functions. To this end, we investigated the effect of introducing females into an otherwise male audience during repeated performances of a short concert performed by an internationally acclaimed male choir, the St. Thomas Choir of Leipzig, famed for its 800-year tradition of excellence under directors including J. S. Bach. Consistent with the dual-function hypothesis, it was expected that the presence of females might elicit vocal embellishments that allow individual males to compete for female attention in a manner that does not undermine ensemble cohesion. Moreover, this effect should be most prevalent in sexually mature males who have already undergone puberty-related maturational changes that deepen the voice, that is, in members of the tenor and bass sections of the choir.

To detect the anticipated vocal embellishments, we recorded individual singers from the choir and then conducted acoustic analyses of spectral energy, sound intensity, performance tempo, and tone onset timing. It was expected that spectral energy (which affects vocal tone quality) would be a likely locus of effects of female presence, as vocal spectrum can be varied without influencing performance parameters such as tuning and timing that could compromise ensemble balance and coordination. Furthermore, aspects of performance that affect ensemble cohesion in terms of timing and balance may vary across repeat performances due to practice effects. To take these effects into account, we measured the degree of tempo matching across parts, interpersonal asynchrony, and balance in terms of relative loudness. Modulations of these indices of ensemble cohesion were not expected to occur due to the extensive experience of the St. Thomas Choir with group performance in general and also with the specific pieces that were sung.

Instead, we were particularly interested in a high frequency band of voice's spectrum, known as the "singer's formant" (∼2,500–3,500 Hz). Increased energy in the singer's formant adds carrying power and brilliance to the voice by emphasizing a frequency region to which human hearing is maximally sensitive (Sundberg, 1974, 1999). In accompanied solo singing, a prominent singer's formant assists the voice to stand out from background instrumental sounds, thus allowing opera singers to project their voices above loud orchestras, which contain relatively little energy in this frequency region (Sundberg, 1987). The use of the singer's formant in choral singing is a matter of debate. Although, enhancement of the singer's formant is not typically advocated in choral singing (where the focus is on the blending of voices; Ternström, 1991, 2003), it can nevertheless be observed in choral performance under some circumstances (especially when the singers are also experienced soloists; Rossing et al., 1986).

# MATERIALS AND METHODS

## Participants: The St. Thomas Choir

Sixteen members of the St. Thomas Choir of Leipzig, a boys' choir in Germany, participated in the study ($N = 16$): Four sopranos, four altos, four tenors, and four basses. Sopranos were aged 12 ($n = 2$) and 13 years ($n = 2$); altos were aged 12 ($n = 2$), 13 ($n = 1$), and 16 years ($n = 1$, singing falsetto); tenors were aged 16 ($n = 2$) and 18 years ($n = 2$); basses were aged 16 ($n = 1$), 17 ($n = 2$), and 19 years ($n = 1$). All singers were naïve to the purpose of the study. A senior member of the choir (a tenor, aged 18 years) who held the position of prefect and was also naïve to the purpose of the study served as the conductor during the recording session.

All participants had thorough musical training, high levels of vocal skill, and extensive experience singing together. The St. Thomas Choir, which is an acclaimed ensemble that rehearses daily, tours globally, produces commercial recordings, and maintains a schedule of three performances per week that attract crowds of tourists to the St. Thomas Church in Leipzig: http://www.thomanerchor.de. The choir was established in 1212 and has run continuously since then under the leadership of esteemed cantors (musical directors), including Johann Sebastian Bach. Members of the choir all attend the St. Thomas School, a public boarding school, from the fourth to twelfth class. The music training that choir members receive during this period includes weekly singing and instrumental lessons, music theory classes, instruction in phonation, and daily individual practice sessions and rehearsal of the choir in full and in sections.

At the time of data collection for the current study, the St. Thomas Choir included 103 members in total. The 16 students who participated in our study were recruited following earlier consultation with the cantor's assistant, who was asked to recommend the best available choristers. Thus, although the participant sample is small, it comprises elite individuals with highly honed vocal and ensemble skills. Participants' parents provided written informed consent for their sons to take part in the study, which was run in accordance with the declaration of Helsinki and protocol approved by the ethics committee of the University of Leipzig.

## Materials

The materials consisted of two distinctive pieces of choral music composed by Johann Sebastian Bach. One (a chorale) requires strict synchrony between soprano, alto, tenor, and bass sections of the choir and the other (a fugue) is characterized by greater

rhythmic independence between voices. Both pieces are in the standard repertoire of the St. Thomas Choir, and all participants were familiar with the pieces and had sung them on numerous previous occasions.

The chorale "Du heilige Brunst, süßer Trost [You holy zeal, sweet consolation]" is from the motet "Der Geist hilft unser Schwachheit auf [The Spirit gives aid to our weakness]" (BWV 226) composed by Bach in Leipzig in 1729. The chorale setting is 24 bars long, in quadruple meter (with 4 quarter-note beats per bar), homophonic in texture (although parts contain passing notes), and, with its stately rhythm and holy text (from a Pentecostal hymn written by Martin Luther), may be described as "reverent" in character. The fugue comes from the motet "Singet dem Herrn ein neues Lied [Sing unto the Lord a new song]" (BWV 225), first performed in Leipzig around 1727. The fugue—a setting of biblical text "Alles, was Odem hat, lobe den Herrn [All that have voice, praise the Lord]" from Psalm 150:6—spans 112 bars (from bar 256 to the end of the work), is in triple meter (with 3 eighth-note beats per bar), polyphonic in texture, and has a "lively" character featuring rapid runs of sixteenth notes cascading through the choral parts.

Both pieces are written in the key of B-flat major with four vocal parts: soprano, alto, tenor, and bass. For the chorale, the soprano part contains 95 notes ranging in pitch from F4 to G5 (349–784 Hz), the alto part has 104 notes ranging from C4 to D5 (262–587 Hz), the tenor part has 110 notes ranging from E-flat3 to F4 (156–349 Hz), and the bass part has 120 notes ranging from B-flat2 to D4 (117–294 Hz). For the fugue, the soprano part has 320 notes (F4 to B-flat5; 349–932 Hz), the alto part has 334 notes (B-flat3 to D5; 233–587 Hz), the tenor part has 369 notes (F3 to A5; 175–440 Hz), and the bass part has 419 notes (G2 to E-flat4; 98–311 Hz).

## Procedure

The study took place in the chamber music auditorium at Castle Colditz in Germany, where the St. Thomas Choir was interned at their annual summer camp. The auditorium contained a stage area, which was large enough to accommodate the choir comfortably, and six rows of tiered seating. The procedure was completed in a single 1-h session. Prior to commencing the session, the 16 members of the choir and the conductor were assembled and introduced to the researchers. They were told that the researchers were studying choral singing and were specifically interested in the level of balance, timing precision, and intonation achieved in a top choir. Members of the choir and the conductor were given the musical scores for the two pieces and were told that the upcoming task involved singing three repeat performances of the pieces.

Choir members were then fitted with individual head-worn wireless microphones (Sennheiser ME 3). Signals picked up by the microphones were sent via radio transmitters (Sennheiser SK 500) to a receiver (Sennheiser EW 500) connected to a stage unit (Roland S-1608) with built-in analog-to-digital converter and microphone preamplifiers. The stage unit routed the audio to a digital mixer (Roland M-300), which was used to record each singer's voice onto a separate channel with digital audio software (Cakewalk Sonar X1 Producer) running on a laptop computer (Dell Inspiron with REAC driver). The equipment was configured to record audio in each channel in mono at a sampling rate of 48 kHz in .WAV format. In addition to this multi-track audio recording, two digital video cameras (Sony HDR-HC7) were used to record the performances. One camera was positioned at the rear of the auditorium and was oriented in such a way that it captured a frontal view of the choir. The second camera, which was placed to the side of the choir, recorded a frontal view of the audience.

Once the audio set up was complete, choir members were asked to stand in two rows, as per their conventional formation, with the front row consisting of the four sopranos on the left and the four altos on the right, and the back row consisting of the four tenors on the left (behind the sopranos) and the four basses on the right (behind the altos). The conductor was positioned in front of the choir, facing the singers (and could therefore not see the audience while conducting). When in position, a sound check was performed to verify that each channel was receiving a clear signal and to optimize recording levels. The conductor was then asked to run through each piece once with the choir in preparation for the recording session.

The choir was informed that there would be researchers around throughout the recording session, as well as a small audience to create a natural performance situation. The choir was also told that a "school group" was due to arrive for a castle tour and would join the audience for a performance and would be taking notes for a school magazine article. This school group consisted of four females who constituted our manipulation (described below). The choir was instructed to sing as they normally would in a concert performance.

In the recording session, the choir performed the two pieces, one after the other, three times each in a separate "concert." Recordings of the performances are provided as Supplementary Audio. The chorale was performed before the fugue in each concert. Each concert was separated by an intermission lasting 1–3 min, during which the composition of the audience was changed. The first of the three concerts was sung to a small audience comprised of adult male listeners. Four adolescent females were added to the audience for the second concert (which also included the males). The females then departed before the third concert, which was sung to the original male audience. (While we would have liked to vary the type of additional audience members systematically in a larger study design, including younger and older females and males of corresponding ages, such extra manipulations were precluded by the limited availability of the St. Thomas Choir).

The male audience members were four staff members of the St. Thomas School (aged between 35 and 55 years). Also present in the auditorium were two adult male researchers (aged 28 and 42 years), two adult female research assistants (aged 26 and 24 years), and an adult male recording engineer (aged 27 years) and photographer (aged 28 years). The adolescent female audience members were aged 15 ($n = 2$) and 16 years ($n = 2$). These females were relatives of one of the female research assistants, who invited them to attend the recording session. The females were naïve with regard to the goals of the study. Prior to entering the auditorium, each female was given a notepad and

pen and instructed to make a written record of their impressions of the choir's performance during the concert. This task was intended to occupy the females and decrease the likelihood that they would engage in behavior that could possibly distract the choir. Examining the video recordings revealed no such behavior during the performances.

At the conclusion of the recording session, each member of the choir completed a questionnaire to evaluate his subjective impressions of the performances and views on choral singing. Questionnaire items addressed (1) the level of concentration achieved while singing, (2) the focus of attention (on self, neighboring singers, whole choir, conductor, and audience), (3) the degree to which the audience is focused upon generally in concerts, (4) the degree to which the audience proved distracting during the recording session, (5) whether the presence of girls in the audience led to changes in the individual's own performance, and (6) which of the three concerts went the best (see **Appendix**). Item 5, on effects of the girls, included several options that could be selected, including one that directly probed whether the respondent attempted to attract the girls' attention. The conductor was asked which performance he thought went the best.

## Data Analysis

Prior to analysis, the audio recordings of individual voices for each performance of each piece were extracted from the continuous recordings of the full session. The extraction process involved opening the recordings of individual voices simultaneously in separate channels in Cubase LE 6 digital audio software, and then cutting the 16 tracks in such a way that they all started 1 s prior to the onset of the first vocal sound and ended 1 s after the offset of the final vocal sound in each performance for each piece. These individual voice recordings, thus synchronized across choir members, were then subject to acoustic analysis using MIRtoolbox 1.6.1 (Lartillot et al., 2008), a Matlab toolbox for the extraction of musical features from audio files. The following features were analyzed: energy in the singer's formant spectral region (2,500–3,500 Hz), sound intensity (to assess ensemble balance), global performance tempo, variability in local tempo, and relative note-onset timing (to assess the accuracy and precision of ensemble coordination). This latter analysis also made use of a full choir version for each performance created by mixing the 16 individual voice channels into a single mono file in .WAV format. The feature extraction process is described below.

### Spectral Energy

For data visualization, the spectral envelope was extracted from each individual voice file using the "mirspectrum" function of MIRtoolbox to decompose the signal into its frequency components via Fast Fourier Transform, and then using the "mirenvelope" function to fit a curve tracing the amplitude of each frequency component of this spectrum. The "Terhardt" option was used with "mirspectrum" to modulate the energy following the Terhardt (1979) outer ear model, which emphasizes frequencies around 2,000–5,000 Hz, at which human hearing is particularly sensitive.

For statistical analysis, the proportion of total energy in the 2,500–3,500 Hz spectral region of each individual voice file was computed using the "mirbrightness" function of MIRtoolbox. This function calculates the proportion of energy (0–1) above a user-specified cut-off frequency. We applied the "mirbrightness" function twice to each file, once with a cut-off frequency of 2,500 Hz and then with a cut-off frequency of 3,500 Hz, and subsequently computed the difference to yield the proportion of total energy between 2,500 and 3,500 Hz. A proportional measure was chosen because the relative amount of energy in high frequency regions is a reliable correlate of perceived timbre in vocal expression (von Bismarck, 1974; Laukka et al., 2005).

### Sound Intensity

Global sound intensity was measured by calculating the root mean square of the amplitude profile of each individual voice file using the "mirrms" function from MIRtoolbox. This analysis takes into account energy at all frequencies of the sound spectra, and therefore allows overall ensemble balance in terms of relative intensity of different voice parts to be assessed. The ANOVAs on these data allowed us to test whether female presence affected ensemble balance by changing intensity in some voices relative to others. This could occur if, for example, distraction caused by female presence disturbed the so-called "self-to-other ratio," which reflects the degree to which an individual can hear their own sounds amongst co-performers' sounds (Ternström, 2003).

### Global and Local Tempo

Global tempo was estimated for each individual voice file using the "mirtempo" function in MIRtoolbox. This function detects note onsets based on the amplitude envelope of the audio signal, then searches for periodicities in the note-onset time series for a range of tempi (40–200 beats-per-minute [bpm]), and finally selects the tempo yielding the maximum periodicity score. Variability in local tempo was assessed by calculating the standard deviation of time points in the tempo map for each file. The "Metre" option of "mirtempo" was employed in this analysis to take into account variability at metrical levels other than the beat period (Lartillot et al., 2013).

### Relative Note-onset Timing

Note-onset times were estimated for each individual voice file, as well as for the mixed full choir file for each performance, using the "mironsets" function in MIRtoolbox. This function applies a peak-picking algorithm to an onset detection curve computed based on the audio signal's amplitude envelope. The "SpectroFrame" option was used in this analysis, with the frame length and hop factor both set to 150 ms to avoid the detection of spurious onsets. This value was chosen based on the assumption that, given the notated rhythms in the pieces and the tempi at which they were performed, successive note onsets would be separated by at least 150 ms.

To assess ensemble coordination, asynchronies between note onsets in each individual voice file and onsets in the mixed full choir file were computed for each performance. These asynchronies represent temporal deviations of estimated note onsets produced by each singer from estimated onsets in the

choir's collectively produced sound, which served as a temporal frame of reference for the current analysis. Asynchrony series were generated with an algorithm that, for each note onset in a given individual voice file, searched for the nearest onset in the mixed full choir file and, if these two onsets were separated by <150 ms, recorded the asynchrony. Median asynchrony was taken as a measure of synchronization accuracy and the standard deviation of the asynchronies series was taken as a measure of synchronization precision for each asynchrony series.

### Statistical Tests

Data for spectral energy were entered into a (2 × 2) × 4 mixed Analysis of Variance (ANOVA) with female presence (present, absent) and piece (chorale, fugue) as within-participant factors, and voice type (bass, tenor, alto, soprano) as a between-participants factor. The criterion for statistical significance was set at $\alpha = 0.05$ in all analyses reported here. Significant interaction effects were followed up with two-tailed paired-samples *t*-tests comparing spectral energy when females were present vs. absent for each voice type separately. Given the stylistic differences between the two pieces, affecting musical texture, dynamics, rhythm, and tempo, data for each of the remaining features were entered into a separate (2) × 4 ANOVA, with factors female presence (present vs. absent) and voice type (bass, tenor, alto, soprano), for each piece.

## RESULTS AND DISCUSSION

The results reported below address the effects of the presence of female audience members on individual voices recorded from the soprano, alto, tenor, and bass sections of the St. Thomas Choir during performances of two distinctive musical pieces. Results for spectral energy in individual boys' voices (focusing on 2,500–3,500 Hz "singer's formant" region) are presented first, followed by sound intensity (to assess ensemble balance), global performance tempo, variability in local tempo, and relative tone-onset timing (to assess the accuracy and precision of ensemble coordination).

### Spectral Energy

Spectral energy was computed for individual singers from audio signals for each of performance of each piece. Time-averaged mean spectral magnitude for sopranos, altos, tenors, and basses (4 singers per part), averaged across the two pieces (chorale and fugue), for performances with female audience members present vs. absent are presented in **Figure 1**. For display, the plot shows output of the Terhardt (1979) outer ear model for audio signal inputs at frequencies up to 4,000 Hz. Here it can be seen that there is a relatively high peak in energy in a high frequency band (~2,500–3,500 Hz, marked by vertical dotted lines) of the vocal spectrum, corresponding to the singer's formant, in the bass section.

For statistical analysis, the proportion of total energy in the 2,500–3,500 Hz singer's formant region of each individual voice file was calculated (on the raw audio signal, without applying the Terhardt model). **Figure 2A** shows the proportion of energy in this region for individual singers in each voice section for the three performances of the two pieces. Performance 1 was sung to the male audience, performance 2 to the audience with additional females, and performance 3 to the male audience again. For both pieces, there is an increase in energy from performance 1 to 2 (when females were present), followed by a decrease in performance 3, for each of the four basses, but no such consistent pattern in other voice sections. **Figure 2B** highlights the main result by displaying the proportion of total energy in the singer's formant region for each voice part (averaged across individual singers and pieces) when females were present vs. absent.

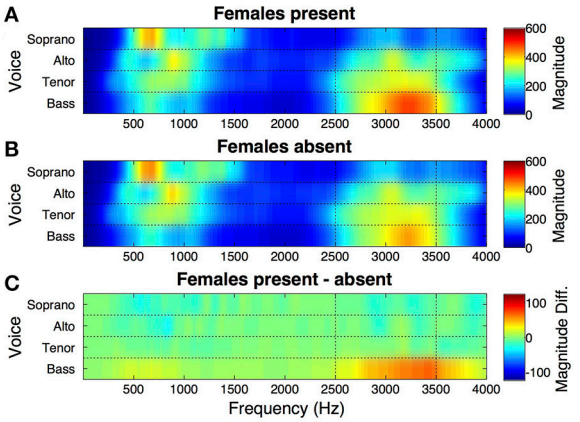



**FIGURE 1 |** Effect of female presence on spectral energy in male voices singing in chorus. **(A)** Time-averaged spectra for choir voice sections singing with adolescent females present in the audience. Output of the Terhardt outer ear model (see Section Spectral Energy) is shown for audio signal inputs at frequencies up to 4,000 Hz in sopranos, altos, tenors, and basses. Audio was recorded with individual head-worn microphones as the choir performed two distinctive pieces. Mean spectral magnitude is shown for each voice section (4 singers per section), averaged across the two pieces. There is a relatively high peak in energy in the singer's formant region (2,500–3,500 Hz) for basses. **(B)** Average spectra for voices when singing the same pieces to a male audience with females absent. The energy peak in basses in the singer's formant region is weaker than when females were present. **(C)** The difference between spectral magnitudes when females were present vs. absent, highlighting the singer's formant increase in basses.

An ANOVA testing for effects of female presence (present, absent), piece (chorale, fugue), and voice type (bass, tenor, alto, soprano) on proportion of energy in the singer's formant range revealed statistically significant main effects of female presence [$F_{(1, 12)} = 15.191$, $p = 0.002$], piece [$F_{(1, 12)} = 25.180$, $p < 0.001$], and voice type [$F_{(3, 12)} = 5.634$, $p = 0.012$], as well as a significant interaction between female presence and voice type [$F_{(3, 12)} = 8.440$, $p = 0.003$]. The remaining interaction effects were not statistically significant ($ps > 0.05$).

The main effect of female presence indicates that energy in the 2,500–3,500 Hz singer's formant region was generally higher when females were present than absent. This effect was, however, qualified by the interaction between female presence and voice type, suggesting that—as can be seen in **Figure 2B**—the bass section was primarily responsible the increase in energy observed when females were present. Paired-samples *t*-tests confirmed that the effect of female presence was statistically significant for the basses [$t_{(3)} = 4.917$, $p = 0.016$] while reliable effects were not

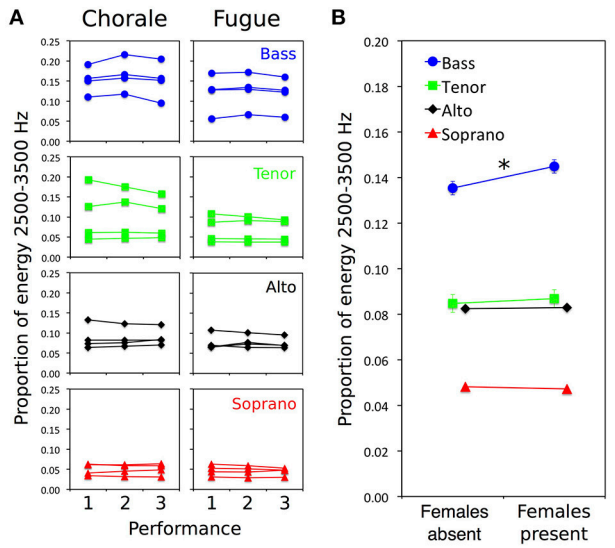



**FIGURE 2 |** Effects of female presence on spectral energy (based on audio signals) in the voices of individual male singers for two distinctive pieces of choir music. **(A)** Proportion of energy in the singer's formant region for individual singers in each voice section for three performances of each piece, a chorale (left) requiring rhythmic synchrony between choir sections and a fugue (right) requiring rhythmic independence between sections. Performance 1 was sung to the male audience, performance 2 to the audience with additional females, and performance 3 to the male audience again. Separate plots are shown for basses (blue circles), tenors (green squares), altos (black diamonds), and sopranos (red triangles). For both pieces, there is an increase in energy from performance 1 to 2 (when females present), followed by a decrease in performance 3, for each of the four basses, but no such consistent effect in other voice sections. **(B)** Proportion of total energy in the singer's formant region for each voice part (averaged across individual singers and pieces) when females were present vs. absent. Proportion of energy is higher when females were present than absent only in basses (female presence × voice type interaction, $p = 0.003$; asterisk indicates $p = 0.016$ for simple effect of female presence for basses only). Error bars are 95% confidence intervals based on within-participants s.e.m. (bars occluded by data marker symbols for sopranos and altos).

observed for tenors ($p = 0.462$), altos ($p = 0.487$), or sopranos ($p = 0.554$).

The specificity of the effect to members of the bass section, that is, to singers with the deepest voices, is noteworthy. A prominent singer's formant is conventionally considered to be more desirable in solo than choral singing (Ternström, 2003). However, because enhancing the singer's formant does not affect other performance parameters such as tuning and timing, the observed effect may reflect sexually mature males competing for female attention in a manner that does not threaten group cohesion.

The main effect of piece indicates that proportion of energy in the 2,500–3,500 Hz range was generally greater for the chorale than the fugue, which may be due to differences in compositional texture and rhythmic structure between the pieces. Specifically, enhancement of the singer's formant might have been elicited by the homogenous texture of the chorale, where separate voice sections sing predominantly in rhythmic unison, more so than for the polyphonic fugue, where the vocal lines sung by separate voice sections are differentiated in terms of rhythm. Furthermore, longer note durations in the chorale (mainly quarter and eighth notes at a steady tempo) than the fugue (eighth and sixteenth notes at a rapid tempo) may generally allow more scope for modulating spectral properties of vocal sound.

The main effect of voice type indicates that singer's formant energy increased across choir sections from sopranos through altos and tenors to basses. This increase may be due in part to differences in the perceptual salience of separate voice parts, as well as to changes in vocal tract size and shape associated with maturational development. With regard to perceptual salience, the general tendency for superior processing of high voices in multi-part musical textures (Palmer and Holleran, 1994; Crawley et al., 2002; Trainor et al., 2014) implies little need to employ additional techniques to enhance the soprano part. With regard to vocal tract configuration, the lowering of the larynx and widening of the pharynx during adolescent development (Fitch and Giedd, 1999) increases the size of a resonance cavity that causes clustering of high formants in the voice spectrum (Sundberg, 1987). Boys who have undergone growth associated with puberty are therefore more likely to display a prominent singer's formant and, in adults, enhancement of the signer's formant is more commonly observed in male tenors and basses than in female sopranos and altos (Sundberg, 1999). Given that participant age was related to voice type (though note that tenors and basses were similar in age), it is not surprising that energy in the 2,500–3,500 Hz region increased across voice types.

It can also be noted that age differences were not responsible for the presence of the effect in basses and absence of the effect in tenors. Although, two tenors exhibited an enhanced singer's formant during the second performance (see **Figure 2A**), these individuals were in fact younger (aged 16 years) than the other two tenors who did not show an effect (aged 18 years).

## Sound Intensity

Global sound intensity was measured by calculating the root mean square of the amplitude profile of each individual voice file. Average data for each voice section for each performance of each piece are shown in **Table 1**. Separate ANOVAs, testing for effects of female presence (present vs. absent) and voice type (bass, tenor, alto, soprano) on sound intensity, for each piece yielded no statistically significant effects ($ps > 0.05$). The absence of significant interactions involving voice type suggests that ensemble balance, in terms of relative intensity of different voice parts, was consistent across performances.

## Global and Local Tempo

Global tempo was estimated for each individual voice file and data for this measure are displayed in **Table 1**. The average estimated global tempo across voices was 80.34 bpm for the chorale and 161.07 bpm for the fugue. An ANOVA on global tempo data for the chorale revealed a statistically significant main effect of female presence [$F_{(1, 12)} = 5.357$, $p = 0.039$] and a significant interaction of female presence and voice type [$F_{(3, 12)} = 4.009$, $p = 0.034$]. The main effect of voice type was not significant. These results reflect an overall slightly *slower* performance tempo when females were present (81.12 bpm) than

**TABLE 1 |** Measures of sound intensity (root mean square of amplitude profile), global tempo (bpm), local tempo variability (SD of onsets in tempo map), median asynchrony (in ms, relative to note onsets in mixed full choir file), and SD of asynchronies (ms) computed from individual audio files for three performances (females present in audience for performance 2) of two pieces (chorale and fugue).

| Measure | Voice | Chorale | | | | | | Fugue | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Performance | | | | | | Performance | | | | | |
| | | 1 | | 2 | | 3 | | 1 | | 2 | | 3 | |
| Sound intensity | Soprano | 0.249 | (0.056) | 0.231 | (0.045) | 0.246 | (0.057) | 0.242 | (0.051) | 0.242 | (0.047) | 0.258 | (0.056) |
| | Alto | 0.207 | (0.024) | 0.171 | (0.017) | 0.171 | (0.018) | 0.255 | (0.017) | 0.237 | (0.009) | 0.226 | (0.009) |
| | Tenor | 0.184 | (0.019) | 0.180 | (0.019) | 0.169 | (0.015) | 0.254 | (0.019) | 0.256 | (0.018) | 0.248 | (0.017) |
| | Bass | 0.106 | (0.009) | 0.117 | (0.015) | 0.106 | (0.010) | 0.152 | (0.014) | 0.179 | (0.025) | 0.163 | (0.018) |
| Global tempo | Soprano | 83.84 | (1.70) | 80.76 | (1.52) | 84.98 | (1.23) | 161.40 | (1.86) | 162.51 | (0.51) | 161.98 | (0.48) |
| | Alto | 77.66 | (0.85) | 81.21 | (1.29) | 81.00 | (2.68) | 158.91 | (0.98) | 160.85 | (0.92) | 160.05 | (1.20) |
| | Tenor | 75.74 | (2.47) | 77.98 | (0.81) | 81.59 | (1.31) | 158.75 | (1.62) | 160.74 | (0.72) | 158.00 | (1.80) |
| | Bass | 81.97 | (2.24) | 78.31 | (1.43) | 82.22 | (2.90) | 160.32 | (0.58) | 163.63 | (0.88) | 162.23 | (1.06) |
| Local tempo | Soprano | 6.46 | (0.94) | 7.78 | (1.87) | 4.90 | (1.13) | 10.33 | (3.14) | 6.17 | (1.55) | 9.23 | (2.24) |
| | Alto | 20.20 | (6.89) | 10.14 | (2.70) | 10.58 | (4.20) | 24.27 | (10.10) | 19.16 | (6.99) | 13.00 | (1.70) |
| | Tenor | 6.23 | (2.04) | 10.34 | (1.97) | 10.89 | (2.25) | 7.61 | (2.20) | 9.14 | (3.61) | 9.55 | (1.96) |
| | Bass | 4.02 | (0.16) | 14.61 | (11.19) | 9.29 | (2.33) | 13.91 | (5.43) | 8.98 | (2.59) | 10.33 | (5.30) |
| Median asynchrony | Soprano | −5.02 | (4.48) | −0.70 | (4.69) | −7.15 | (6.79) | −3.26 | (9.48) | −2.78 | (8.81) | −7.07 | (8.45) |
| | Alto | −4.75 | (3.96) | −3.21 | (4.60) | 2.30 | (2.43) | −5.84 | (3.11) | −3.94 | (2.52) | −1.88 | (3.08) |
| | Tenor | 7.96 | (4.12) | −0.11 | (2.72) | 0.80 | (3.00) | 0.79 | (5.75) | −0.77 | (4.78) | −1.64 | (5.06) |
| | Bass | 4.15 | (5.58) | 8.19 | (1.83) | 1.44 | (2.59) | −0.96 | (4.34) | 4.01 | (4.31) | −0.89 | (4.28) |
| SD of asynchronies | Soprano | 61.23 | (2.55) | 59.01 | (3.05) | 57.03 | (2.42) | 45.81 | (0.67) | 46.75 | (1.69) | 45.11 | (1.14) |
| | Alto | 59.14 | (1.46) | 57.24 | (1.45) | 56.49 | (0.69) | 47.70 | (1.66) | 46.09 | (1.36) | 48.23 | (1.69) |
| | Tenor | 59.48 | (1.82) | 58.58 | (0.95) | 61.42 | (1.96) | 48.89 | (1.31) | 49.75 | (1.61) | 48.45 | (1.84) |
| | Bass | 59.15 | (2.79) | 60.60 | (1.30) | 58.19 | (2.10) | 53.15 | (0.89) | 51.13 | (1.40) | 49.97 | (1.31) |

*See Data Analysis for details. Values averaged across the four singers in each voice section (soprano, alto, tenor, and bass; s.e.m. in parentheses).*

absent (79.56 bpm), with this effect being more pronounced in sopranos and basses than in altos and tenors (although *t*-tests for each voice type failed to reveal any significant effects). In contrast, an ANOVA on tempo data for the fugue only yielded a significant main effect of female presence [$F_{(1, 12)} = 13.453$, $p = 0.003$], reflecting overall slightly *faster* performance tempo when females were present (161.93 bpm) than absent (160.20 bpm). The effects of female presence on global tempo were therefore weak and inconsistent across the two pieces. Variability in local tempo was assessed by calculating the standard deviation of time points in a tempo map for each file. The ANOVAs on local tempo variability data (**Table 1**) yielded no statistically significant effects for either piece ($p$s $> 0.05$).

## Ensemble Coordination

To assess ensemble coordination, asynchronies between note onsets in each individual voice file and onsets in the mixed full choir file were computed for each performance. These asynchronies represent temporal deviations of estimated note onsets produced by each singer from estimated onsets in the choir's collectively produced sound. The median of each individual singer's asynchrony series was calculated as a measure of his synchronization accuracy and the standard deviation of the asynchrony series was taken as a measure of synchronization precision (**Table 1**; Rasch, 1988).

The sign of the median asynchrony (negative or positive) is informative about whether note onsets in a particular voice part occur relatively early or late. In the current performances, the sopranos showed a numerical tendency to lead (median asynchrony = −4.33 ms, averaged across the four singers), followed by altos (−2.89 ms), tenors (1.17 ms), and basses (2.66 ms). These differences were, however, not statistically significant. ANOVAs on synchronization accuracy did not reveal significant effects for either piece ($p$s $> 0.05$). With regard to synchronization precision, the mean standard deviation of asynchronies, averaged across all singers, was 59 ms for the chorale and 48 ms for the fugue, which is commensurate with levels typically observed in expert ensembles (Keller, 2014). ANOVAs on synchronization precision did not yield statistically significant effects ($p$s $> 0.05$), apart from a main effect of voice type in the fugue [$F_{(3, 12)} = 3.961$, $p = 0.036$], reflecting a general decrease in precision from sopranos to basses (which may be a consequence of the increase in number of notes across these parts; see Section Materials).

## Subjective Evaluations

After the recording session, each member of the choir completed a questionnaire addressing his evaluation of the performances and views on choral singing in general. Analysis of the questionnaire items revealed no reliable differences between

voice types in terms of (1) self-reported level of concentration, (2) focus of attention on self, neighboring singers, whole choir, conductor, or audience, (3) general tendency to focus on the audience during concerts, (4) the degree to which the audience proved distracting during the recording, (5) whether the presence of girls led to changes in the individual's own performance. For this last item, none of the boys selected the response option indicating that they attempted to attract the girls' attention. In response to a final item probing views on which of the three concerts went the best, all tenors and basses thought that the one sung in the presence of females was best, while only one out of the four sopranos and no altos thought that this was the case. The conductor (who could not see the audience while conducting) indicated that the performance with females present went the best.

## GENERAL DISCUSSION

The present case study tested the hypothesis—motivated by research on non-human "chorusing" displays (such as those produced by congregations of male insects and frogs to attract females)—that group music making serves cooperative and competitive functions simultaneously. In support of this dual-function hypothesis, acoustic analyses of individual voices from the St. Thomas Choir of Leipzig revealed increased energy in the singer's formant spectral region (2,500–3,500 Hz) in members of the bass section, that is, boys with the deepest voices, when females were included in an otherwise male audience during repeated performances of a concert. This selective effect of female presence generalized across different musical pieces—a chorale requiring synchrony between sections of the choir and a fugue characterized by greater rhythmic independence between voices—and did not disrupt ensemble cohesion in terms of balance or temporal coordination. Enhancing the singer's formant generally modifies vocal tone quality by imbuing the voice with what Helmholtz (1875, p. 116) described as "a clear tinkling of little bells". Observing such subtle embellishment in the context of choral singing suggests that, under conditions where group cohesion is strongly favored, sexually mature males have at their disposal a covert means to compete for female attention.

Evidence that the enhanced singer's formant was motivated by sexual competition lies in the finding that the effect of female presence was only reliably observed in members of the bass section of the choir. The presence of adolescent females (aged 15–16 years) may have elicited competitive behavior exclusively in bass singers (aged 16–19 years) due to relatively high levels of the sex hormone testosterone in males with the deepest voices (Zitzmann and Nieschlag, 2001). Testosterone lengthens the vocal tract by stimulating a secondary descent of the larynx in males during puberty (Fant, 1975). This lengthening lowers the fundamental frequency of the voice and reduces dispersion in formant frequencies (Bruckert et al., 2006), which increases perceived attractiveness and dominance (Evans et al., 2008). The boost in the singer's formant that we observed may constitute an attempt by the basses to establish a privileged social communication channel with female listeners by drawing attention to these appealing vocal qualities.

The fact that female presence did not lead to a reliable enhancement of the singer's formant in tenors (aged 16–18 years) suggests that the effect was not related to a general desire to lift performance standard in the presence of female peers. An increase in the desire to perform well could also be expected to affect other performance parameters, even in sopranos (12–16 years) and altos (12–16 years), but no such additional effects were observed. Furthermore, as revealed by post-concert questionnaire responses, members of the tenor and bass section were unanimous in thinking that the performance sung in the presence of females was the best, but none admitted to attempting to attract the girls' attention. Therefore, while tenors and basses converged in terms of their subjective evaluations of the performances, their objectively measured behavior diverged.

Our results point to an analogy between human choral singing and chorusing displays in non-human species (crickets, cicadas, and frogs) where congregations of males produce rhythmically coordinated signals that collectively serve as a beacon to attract female mates. While non-human chorusing appears cooperative to the extent that inter-individual coordination maximizes the intensity and geographic reach of the collective broadcast, these communal displays can arise via competitive mechanisms through which individual males jam rival signals (Greenfield, 1994, 2005). By extension, the selective enhancement of the singer's formant observed in our study suggests that, in human music, an individual beacon motivated by competition can safely nest within a group beacon arising through cooperation.

Apart from enhancing the singer's formant in basses, female presence did not otherwise impact reliably upon performance. Features including balance in terms of overall vocal intensity, global tempo, local tempo fluctuations, and degree of temporal coordination between voices were generally commensurate across performances with and without female audience members. This indicates that individual singers did not attempt to attract attention by obvious means such as producing louder or earlier sounds than their co-performers—unlike the case in non-human chorusing. Clearly, the tradition of excellence and discipline in the St. Thomas Choir and the presence of a conductor (who was blind to the manipulation) precluded variations in rhythmic precision or vocal blend that would compromise ensemble cohesion. Indeed, the fact that we observed any vocal embellishment at all is remarkable given that the music was composed for religious celebration in the Lutheran Protestant tradition, where flamboyance is eschewed (in contrast to opera, gospel, and pop choirs, for example).

The absence of effects for measures of ensemble timing and balance also argues against the possibility that the enhancement of the singer's formant for the second of three performances represents a practice effect. Aspects of performance that are central to the explicit goal of achieving ensemble cohesion, such as timing and balance, are presumably more likely to be susceptible to practice effects than unsanctioned behaviors such as the observed spectral modifications. In any case, practice effects were not expected to occur with the St. Thomas Choir due to their extensive experience with the material performed.

Fatigue effects were likewise not expected due to the choir's routine experience singing much longer programs, including masses and oratorios lasting several hours.

As noted earlier, our approach entails a single case study of an elite musical ensemble rather than a fully controlled experiment addressing general music making in a random sample of individuals. Conducting a case study with experts allowed us to provide a hard test of the hypothesis that female presence selectively impacts upon spectral rather than temporal or intensity aspects of male vocal production, and that this effect is specific to males who have undergone voice changes associated with puberty. These goals would not be readily achieved with alternative study designs. Amateur choirs typically comprise sections in which individuals are mixed in terms of age and experience (especially amongst tenors and basses), making these variables difficult to place under experimental control. Forming an *ad hoc* choir for the purpose of the study would likewise be problematic due to the fact that new ensembles usually spend considerable time—several weeks of regular rehearsal—learning to produce a cohesive sound. Moreover, true ensemble excellence is a hallmark of years of working together (Murnighan and Conlon, 1991; Blank and Davidson, 2007). A choir lacking such extensive preparation would most likely produce a high degree of variation in performance timing and balance, which could mask subtle spectral effects associated with the singer's formant.

Of course, investigating a single elite group has the obvious limitation of a small sample size. For this reason, caution should be taken in drawing generalized conclusions, especially since the current study focused on a single, culture-specific musical style. Indeed, one may ask whether effects observed with a highly drilled European boys choir would generalize to everyday forms of group music making in different cultures. While we cannot answer this question based on present findings, work in the field of ethnomusicology suggests that such generalizability is feasible. Group singing is vital to the existence of many indigenous cultures. For example, the male Amazonian Mekranoti Indians sing daily before dawn in a defensive vigil that maintains arousal to guard against attacks from neighboring tribes (Werner, 1984; Huron, 2001) while the nearby Suya people use song as a routine communicative device that complements and overlaps with speech (Seeger, 1987). It is possible that acoustic analyses of these everyday group vocal displays would reveal spectral modulations based on social communicative goals.

Notwithstanding caveats related to the case study approach, our results have several implications for future studies aimed at understanding human non-verbal communication from the perspectives of evolutionary biology, psychology, music cognition, and bioacoustics. With regard to the evolutionary biology, the concept of simultaneous cooperation and competition points to a relationship between human and non-human "chorusing" that goes beyond the fact that both involve rhythmic coordination. Building on this link, the hypothesis that the evolutionary benefit of music lies in balancing sexually motivated behavior and group cohesion advances theory by bringing together opposing accounts of the origins of music, thus highlighting a promising avenue for cross-species comparative research. With regard to psychology, our study highlights music's potential to advance knowledge about the dynamics of non-verbal communication during social interaction. In particular, demonstrating how cooperative and competitive mechanisms operate in parallel—when differing goals are pursued at the level of the group and the individual—showcases music as an ecologically valid domain in which to investigate this social balancing act under controlled conditions (D'Ausilio et al., 2015).

Finally, with regard to music cognition and bioacoustics, the current study is informative about the role of sound quality, in particular the singer's formant, in communicating performers' intentions. In addition to its well-documented function in adding carrying power to the voice in operatic soloists (Sundberg, 1987), our results suggest that the signer's formant may be used to draw the listener's attention to a particular auditory stream in a multipart choral texture. It is important to note that this effect was observed in an elite ensemble under conditions that ostensibly favor the blending of voices. This emphasizes the need for models of musical expression, which to date focus predominantly on variations of timing and intensity (Todd, 1985, 1992; Clarke, 1988; Palmer, 1997; Gabrielsson, 2003; Keller, 2012), to take into account the role of spectral properties in signaling performers' intentions in ensemble music.

In conclusion, the findings of the present study suggest that, analogously to non-human chorusing, human music making may enable the coexistence of cooperative and competitive behavior that potentially mediates social communication simultaneously at the level of the group and the individual. Even during highly formalized types of musical interaction (exemplified here by Western church music), performers are free to introduce unsanctioned behavioral embellishments that allow individuals to enter into covert competition for the attention of potential mates without undermining collaborative goals related to emotional communication and artistic expression. The fact that we observed such subtle modifications in the singing voice—an ancient and universal means of musical expression (Lomax and Berkowitz, 1972; Nettl, 1983; Mithen, 2005; Hagen and Hammerstein, 2009)—raises the possibility that balancing sexual competition with group cohesion in musical contexts is a human capacity with deep biological roots. Accounts of the origins of human music in social cohesion vs. sexual selection can therefore be unified in the hypothesis that the evolutionary benefit of music lies in its ability to serve cooperative and competitive functions simultaneously, with this duality augmenting music's communicative power.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ethics committee of the University of Leipzig with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committee of the University of Leipzig.

## AUTHOR CONTRIBUTIONS

RK and PK designed the study with input from GN. RK and PK collected and analyzed the data. All authors discussed the results. PK wrote the manuscript with input from RK and GN.

## FUNDING

Funding from the Max Planck Society supported the study. PK is supported by a Future Fellowship grant from the Australian Research Council (FT140101162).

## ACKNOWLEDGMENTS

We thank the members of the St. Thomas Choir, their management team, including Roland Weise, Thoralf Schulze, and Titus Heidemann, and former cantor Georg Christoph Biller for making the study possible. We also thank Christoph Wonneberger, Marko Kronberg, Markus Eckardt, and Stephan Liebig for technical assistance, and Maria Bader and Theres König for assistance with data collection.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpsyg.2017.01559/full#supplementary-material

Supplementary material includes audio recordings of the full choir's three performances of the two pieces:

**Audio S1 |** chorale_1_females_absent.mp3

**Audio S2 |** chorale_2_females_present.mp3

**Audio S3 |** chorale_3_females_absent.mp3

**Audio S4 |** fugue_1_females_absent.mp3

**Audio S5 |** fugue_2_females_absent.mp3

**Audio S6 |** fugue_3_females_absent.mp3

Note that the 16 separate channels (one per choir member) are mixed in these files to give an overall impression of the choir's performances, and the files have been compressed for presentation purposes. Uncompressed files are available upon request from PK (p.keller@westernsydney.edu.au).

## REFERENCES

Bailey, W. J. (1991). *Acoustic Behaviour of Insects. An Evolutionary Perspective*. London: Chapman and Hall.

Blacking, J. (1973). *How Musical is Man*. Seattle, WA: University of Washington Press.

Blank, M., and Davidson, J. W. (2007). An exploration of the effects of musical and social factors in piano duo collaborations. *Psychol. Music* 35, 231–248. doi: 10.1177/0305735607070306

Branham, M. A., and Greenfield, M. D. (1996). Flashing males win mate success. *Nature* 381, 745–746. doi: 10.1038/381745b0

Brown, S. (2000). "Evolutionary models of music: from sexual selection to group selection," in *Perspectives in Ethology: Evolution, Culture, and Behavior*, eds F. Tonneau and N. S. Thompson (Boston, MA: Springer), 231–281.

Bruckert, L., Lienard, J. S., Lacroix, A., Kreutzer, M., and Leboucher, G. (2006). Women use voice parameters to assess men's characteristics. *Proc. R. Soc. Lond. B Biol. Sci.* 273, 83–89. doi: 10.1098/rspb.2005.3265

Buck, J. (1988). Synchronous rhythmic flashing of fireflies. II. *Q. Rev. Biol.* 63, 265–289. doi: 10.1086/415929

Buck, J., and Buck, E. (1966). Biology of synchronous flashing of fireflies. *Nature* 211, 562–564. doi: 10.1038/211562a0

Catchpole, C. K. (1983). Variation in the song of the great reed warbler Acrocephalus arundinaceus in relation to mate attraction and territorial defence. *Anim. Behav.* 31, 1217–1225. doi: 10.1016/S0003-3472(83)80028-1

Catchpole, C. K. (1986). Song repertoires and reproductive success in the great reed warbler Acrocephalus arundinaceus. *Behav. Ecol. Sociobiol.* 19, 439–445. doi: 10.1007/BF00300547

Catchpole, C., and Slater, P. (1995). *Bird Song: Biological Themes and Variations*. Cambridge: Cambridge University Press.

Catchpole, C., Leisler, B., and Dittami, J. (1986). Sexual differences in the responses of captive great reed warblers (*Acrocephalus arundinaceus*) to variation in song structure and repertoire size. *Ethology* 73, 69–77. doi: 10.1111/j.1439-0310.1986.tb01000.x

Chaverri, G., Gillam, E. H., and Kunz, T. H. (2012). A call-and-response system facilitates group cohesion among disc-winged bats. *Behav. Ecol.* 24, 481–487. doi: 10.1093/beheco/ars188

Clarke, E. F. (1988). "Generative principles in music performance," in *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition,* ed J. A. Sloboda (New York, NY: Claredon Press), 1–26.

Clayton, M. (2009). "The social and personal functions of music in cross-cultural perspective," in *The Oxford Handbook of Music Psychology,* eds S. Hallam, I. Cross, and M. Thaut (Oxford: Oxford University Press), 35–44.

Clayton, M., Herbert, T., and Middleton, R. (eds.). (2012). *The Cultural Study of Music: A Critical Introduction, 2nd Edn*. London: Routledge.

Crawley, E. J., Acker-Mills, B. E., Pastore, R. E., and Weil, S. (2002). Change detection in multi-voice music: the role of musical structure, musical training, and task demands. *J. Exp. Psychol.* 28, 367–378. doi: 10.1037/0096-1523.28.2.367

Cross, I. (2012). "Music and biocultural evolution," in *The Cultural Study of Music: a Critical Introduction, 2nd Edn*, eds M. Clayton, T. Herbert, and R. Middleton, (London: Routledge), 17–27.

Cross, I., and Morley, I. (2009). "The evolution of music: theories, definitions and the nature of the evidence," in *Communicative Musicality: Exploring the Basis of Human Companionship,* eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 61–81.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

D'Ausilio, A., Novembre, G., Fadiga, L., and Keller, P. E. (2015). What can music tell us about social interaction? *Trends Cogn. Sci.* 19, 111–114. doi: 10.1016/j.tics.2015.01.005

Eriksson, D., and Wallin, L. (1986). Male bird song attracts females—a field experiment. *Behav. Ecol. Sociobiol.* 19, 297–299. doi: 10.1007/BF00300645

Evans, S., Neave, N., Wakelin, D., and Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiol. Behav.* 93, 783–788. doi: 10.1016/j.physbeh.2007.11.033

Fant, G. (1975). Non-uniform vowel normalization. *Speech Transm. Lab. Q. Prog. Status Rep.* 16, 1–19.

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009

Fitch, W. T., and Giedd, J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J. Acoust. Soc. Am.* 106, 1511–1522. doi: 10.1121/1.427148

Gabrielsson, A. L. F. (2003). Music performance research at the millennium. *Psychol. Music* 31, 221–272. doi: 10.1177/03057356030313002

Geissmann, T. (2000). "Gibbon song and human music from an evolutionary perspective," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 103–123.

Greenfield, M. D. (1994). Cooperation and conflict in the evolution of signal interactions. *Annu. Rev. Ecol. Syst.* 25, 97–126. doi: 10.1146/annurev.es.25.110194.000525

Greenfield, M. D. (2005). Mechanisms and evolution of communal sexual displays in arthropods and anurans. *Adv. Study Behav.* 35, 1–62. doi: 10.1016/S0065-3454(05)35001-7

Greenfield, M. D., and Roizen, I. (1993). Katydid synchronous chorusing is an evolutionarily stable outcome of female choice. *Nature* 364, 618–620. doi: 10.1038/364618a0

Greenfield, M. D., Tourtellot, M. K., and Snedden, W. A. (1997). Precedence effects and the evolution of chorusing. *Proc. R. Soc. B Biol. Sci.* 264, 1355–1361. doi: 10.1098/rspb.1997.0188

Hagen, E. H., and Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Hum. Nat.* 14, 21–51. doi: 10.1007/s12110-003-1015-z

Hagen, E. H., and Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Mus. Sci.* 13, 291–320. doi: 10.1177/1029864909013002131

Helmholtz, H. L. F. (1875). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Transl. by A. J. Ellis). London: Longmans, Green, and Co.

Hoppensteadt, F. C., and Keller, J. B. (1976). Synchronization of periodical cicada emergences. *Science* 194, 335–337. doi: 10.1126/science.987617

Huron, D. (2001). Is music an evolutionary adaptation? *Ann. N.Y. Acad. Sci.* 930, 43–61. doi: 10.1111/j.1749-6632.2001.tb05724.x

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (*Tursiops truncatus*). *Science* 289, 1355–1357. doi: 10.1126/science.289.5483.1355

Johnson, L. S., and Searcy, W. A. (1996). Female attraction to male song in house wrens (*Troglodytes aedon*). *Behaviour* 133, 357–366. doi: 10.1163/156853996X00495

Keller, P. E. (2012). "What movement force reveals about cognitive processes in music performance," in *Art in Motion II,* ed A. Mornell (Frankfurt: Peter Lang), 115–153.

Keller, P. E. (2014). "Ensemble performance: interpersonal alignment of musical expression," in *Expressiveness in Music Performance: Empirical Approaches across Styles and Cultures,* eds D. Fabian, R. Timmers, and E. Schubert (Oxford: Oxford University Press), 260–282.

Kirschner, S., and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evol. Hum. Behav.* 31, 354–364. doi: 10.1016/j.evolhumbehav.2010.04.004

Lartillot, O., Cereghetti, D., Eliard, K., Trost, W. J., Rappaz, M.-A., and Grandjean, D. (2013). "Estimating tempo and metrical features by tracking the whole metrical hierarchy," in *Proceedings of the 3rd International Conference on Music and Emotion (ICME3)*, eds G. Luck & O. Brabant (Jyväskylä).

Lartillot, O., Toiviainen, P., and Eerola, T. (2008). "A matlab toolbox for music information retrieval," in *Data Analysis, Machine Learning and Applications,* eds C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker (Berlin: Springer), 261–268.

Laukka, P., Juslin, P., and Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cogn. Emot.* 19, 633–653. doi: 10.1080/02699930441000445

Launay, J., Tarr, B., and Dunbar, R. I. M. (2016). Synchrony as an adaptive mechanism for large-scale human social bonding. *Ethology* 122, 779–789. doi: 10.1111/eth.12528

Locascio, J. V., and Mann, D. A. (2011). Diel and seasonal timing of sound production by black drum (*Pogonias cromis*). *Fish. Bull.* 109, 327–338.

Lomax, A., and Berkowitz, A. L. (1972). The evolutionary taxonomy of culture. *Science* 177, 228–239. doi: 10.1126/science.177.4045.228

Marler, P. (2000). "Origins of music and speech: Insights from animals," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 31–48.

McDermott, J. (2008). The evolution of music. *Nature* 453, 287–288. doi: 10.1038/453287a

McDermott, J., and Hauser, M. (2005). The origins of music: innateness, uniqueness, and evolution. *Music Percept.* 23, 29–59. doi: 10.1525/mp.2005.23.1.29

Merker, B. H. (2000). "Synchronous chorusing and human origins," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge: MA: MIT Press), 315–327.

Merker, B. H., Morley, I., and Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140095. doi: 10.1098/rstb.2014.0095

Merriam, A. P. (1964). *The Anthropology of Music*. Chicago, IL: Northwestern University Press.

Miller, G. (2000). "Evolution of human music through sexual selection," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 329–360.

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld and Nicholson.

Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology, and the Origins of Musicality*. Oxford: Oxford University Press.

Mountjoy, D. J., and Lemon, R. E. (1991). Song as an attractant for male and female European starlings, and the influence of song complexity on their response. *Behav. Ecol. Sociobiol.* 28, 97–100. doi: 10.1007/BF00180986

Murai, M., and Backwell, P. R. Y. (2006). A conspicuous courtship signal in the fiddler crab Uca perplexa: female choice based on display structure. *Behav. Ecol. Sociobiol.* 60, 736–741. doi: 10.1007/s00265-006-0217-x

Murnighan, J. K., and Conlon, D. E. (1991). The dynamics of intense work groups: a study of British string quartets. *Adm. Sci. Q.* 36, 165–186. doi: 10.2307/2393352

Nettl, B. (1983). *The Study of Ethnomusicology: Twenty-Nine Issues and Concepts*. Chicago, IL: University of Illinois Press.

Otte, D., and Loftus-Hills, J. (1979). Chorusing in syrbula (Orthoptera: Acridiae). Cooperation, interference competition, or concealment? *Entomol. News* 90, 159–165.

Palmer, C. (1997). Music performance. *Annu. Rev. Psychol.* 48, 115–138. doi: 10.1146/annurev.psych.48.1.115

Palmer, C., and Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Percept. Psychophys.* 56, 301–312. doi: 10.3758/BF03209764

Partridge, B., and Krebs, J. R. (1978). Tree frog choruses: a mixed evolutionarily stable strategy? *Anim. Behav.* 26, 959–960. doi: 10.1016/0003-3472(78)90160-4

Rasch, R. A. (1988). "Timing and synchronization in ensemble performance," in *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, ed J. A. Sloboda (Oxford: Clarendon Press), 70–90.

Ravignani, A., Bowling, D. L., and Fitch, W. T. (2014). Chorusing, synchrony, and the evolutionary functions of rhythm. *Front. Psychol.* 5:1118. doi: 10.3389/fpsyg.2014.01118

Rossing, T. D., Sundberg, J., and Ternström, S. (1986). Acoustic comparison of voice use in solo and choir singing. *J. Acoust. Soc. Am.* 79, 1975–1981. doi: 10.1121/1.393205

Seeger, A. (1987). *Why Suyá Sing: A Musical Anthropology of an Amazonian People*. New York, NY: Cambridge University Press.

Smith, J. N., Goldizen, A. W., Dunlop, R. A., and Noad, M. J. (2008). Songs of male humpback whales, Megaptera novaeangliae, are involved in intersexual interactions. *Anim. Behav.* 76, 467–477. doi: 10.1016/j.anbehav.2008.02.013

Snedden, W. A., and Greenfield, M. D. (1998). Females prefer leading males: relative call timing and sexual selection in katydid choruses. *Anim. Behav.* 56, 1091–1098. doi: 10.1006/anbe.1998.0871

Sundberg, J. (1974). Articulatory interpretation of the "singing formant". *J. Acoust. Soc. Am.* 55, 838–844. doi: 10.1121/1.1914609

Sundberg, J. (1987). *The Science of the Singing Voice*. Dekalb, IL: Northern Illinois University Press.

Sundberg, J. (1999). "Perception of singing," in *The Psychology of Music, 2nd Edn*, ed D. Deutsch (London: Academic Press), 171–214.

Tarr, B., Launay, J., and Dunbar, R. I. (2014). Music and social bonding: "self-other" merging and neurohormonal mechanisms. *Front. Psychol.* 5:1096. doi: 10.3389/fpsyg.2014.01096

Terhardt, E. (1979). Calculating virtual pitch. *Hear. Res.* 1, 155–182. doi: 10.1016/0378-5955(79)90025-X

Ternström, S. (1991). Physical and acoustic factors that interact with the singer to produce the choral sound. *J. Voice* 5, 128–143. doi: 10.1016/S0892-1997(05)80177-8

Ternström, S. (2003). Choir acoustics-an overview of scientific research published to date. *Int. J. Res. Choral Sing.* 1, 3–12.

Todd, N. P. M. (1985). A model of expressive timing in tonal music. *Music Percept.* 3, 33–57. doi: 10.2307/40285321

Todd, N. P. M. (1992). The dynamics of dynamics: a model of musical expression. *J. Acoust. Soc. Am.* 91, 3540–3550. doi: 10.1121/1.402843

Trainor, L. J., Marie, C., Bruce, I. C., and Bidelman, G. M. (2014). Explaining the high voice superiority effect in polyphonic music: evidence from cortical evoked potentials and peripheral auditory models. *Hear. Res.* 308, 60–70. doi: 10.1016/j.heares.2013.07.014

von Bismarck, G. (1974). Sharpness as an attribute of the timbre of steady sounds. *Acta Acustica United with Acustica*, 30, 159–172.

Walker, T. J. (1969). Acoustic synchrony: two mechanisms in the snowy tree cricket. *Science* 166, 891–894. doi: 10.1126/science.166.3907.891

Weinstein, D., Launay, J., Pearce, E., Dunbar, R. I., and Stewart, L. (2016). Group music performance causes elevated pain thresholds and social bonding in small and large groups of singers. *Evol. Hum. Behav.* 37, 152–158. doi: 10.1016/j.evolhumbehav.2015.10.002

Werner, D. (1984). *Amazon Journey: An Anthropoligist's Year among Brazil's Mekranoti Indians*. New York, NY: Simon & Schuster.

Whitney, C. L., and Krebs, J. R. (1975). Mate selection in Pacific tree frogs. *Nature* 255, 325–326. doi: 10.1038/255325a0

Wiley, R. H., Hatchwell, B. J., and Davies, N. B. (1991). Recognition of individual males' songs by female dunnocks: a mechanism increasing the number of copulatory partners and reproductive success. *Ethology* 88, 145–153. doi: 10.1111/j.1439-0310.1991.tb00269.x

Zitzmann, M., and Nieschlag, E. (2001). Testosterone levels in healthy men and the relation to behavioural and physical characteristics: facts and constructs. *Eur. J. Endocrinol.* 144, 183–197. doi: 10.1530/eje.0.1440183

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Keller, König and Novembre. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

## APPENDIX

Note that this questionnaire has been translated into English from the German original used in the study.

**Questionnaire**

**Please enter your number here:** _____

**Age:** _____

**Voice group:** _____

1. How well were you able to concentrate *today* while singing?

| not at all | | moderately | | strongly |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

2. On who or what do you concentrate generally while singing? (multiple responses permitted)

o Own voice
o Neighbors
o Audience
o Overall choral sound
o Conductor
o Other: _____

3. Do you sometimes concentrate on the audience during concerts?

| not at all | | moderately | | strongly |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

4. Did the audience distract you while singing *today?*

| not at all | | moderately | | strongly |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

5. What changes did you feel in yourself while the school students were present? *(multiple responses permitted)*
   o I tried to concentrate more on my singing
   o I generally concentrated more than usual
   o I was nervous
   o I tried to attract the students' attention
   o I tried to sing louder
   o I did not experience any changes
   o Other: _____

6. In your opinion, which part (block) went best for all of you today?
   o Part 1 o Part 2 o Part 3

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 September 2017
doi: 10.3389/fpsyg.2017.01595

# Neurodiversity, Giftedness, and Aesthetic Perceptual Judgment of Music in Children with Autism

*Nobuo Masataka**

*Primate Research Institute, Kyoto University, Inuyama, Japan*

The author investigated the capability of aesthetic perceptual judgment of music in male children diagnosed with autism spectrum disorder (ASD) when compared to age-matched typically developing (TD) male children. Nineteen boys between 4 and 7 years of age with ASD were compared to 28 TD boys while listening to musical stimuli of different aesthetic levels. The results from two musical experiments using the above participants, are described here. In the first study, responses to a Mozart minuet and a dissonant altered version of the same Mozart minuet were compared. In this first study, the results indicated that both ASD and TD males preferred listening to the original consonant version of the minuet over the altered dissonant version. With the same participants, the second experiment included musical stimuli from four renowned composers: Mozart and Bach's musical works, both considered consonant in their harmonic structure, were compared with music from Schoenberg and Albinoni, two composers who wrote musical works considered exceedingly harmonically dissonant. In the second study, when the stimuli included consonant or dissonant musical stimuli from different composers, the children with ASD showed greater preference for the aesthetic quality of the highly dissonant music compared to the TD children. While children in both of the groups listened to the consonant stimuli of Mozart and Bach music for the same amount of time, the children with ASD listened to the dissonant music of Schoenberg and Albinoni longer than the TD children. As preferring dissonant music is more aesthetically demanding perceptually, these results suggest that ASD male children demonstrate an enhanced capability of aesthetic judgment of music. Subsidiary data collected after the completion of the experiment revealed that absolute pitch ability was prevalent only in the children with ASD, some of whom also possessed extraordinary musical memory. The implications of these results are discussed with reference to the broader notion of neurodiversity, a term coined to capture potentially gifted qualities in individuals diagnosed with ASD.

**Keywords: autism spectrum disorder, music, aesthetic judgment, consonance, neurodiversity, Spielmann (wandering minstrel)**

OPEN ACCESS

***Edited by:***
*Leonid Perlovsky,*
*Harvard University and Air Force Research Laboratory, United States*

***Reviewed by:***
*Robin W. Wilkins,*
*Joint School for Nanoscience and Nanoengineering Gateway MRI Center, United States*
*Alexander Ovsich,*
*Boston College, United States*

***Correspondence:**
*Nobuo Masataka*
*masataka.nobuo.7r@kyoto-u.ac.jp*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *22 April 2017*
***Accepted:*** *31 August 2017*
***Published:*** *22 September 2017*

***Citation:***
*Masataka N (2017) Neurodiversity, Giftedness, and Aesthetic Perceptual Judgment of Music in Children with Autism. Front. Psychol. 8:1595. doi: 10.3389/fpsyg.2017.01595*

## INTRODUCTION

Neurodiversity refers to the notion that seemingly 'impaired' cognitive as well as emotional properties characteristic of developmental disorders such as autism spectrum disorders (ASD) a neurodevelopmental disorder with unusual sensory processing, are not necessarily deficits, but fall into normal behavioral variations exhibited by humans. Stated more formally, this notion was recently described as "a concept where neurological differences are to be recognized and respected as any other human variation" (Armstrong, 2012). The term was first coined in the late 1990s by New York journalist Harvey Blume and Australian autism activist Judy Singer, and has become an important component of the civil rights movement for those with neurologically based disabilities. While this is indeed a paradigm shift from "deficit-oriented view of ASD" to "strength-oriented view of ASD" (Silberman, 2015), experimental and empirical scientific evidence confirming this conceptual notion and paradigm has been meager.

From the archeological perspective, however, a recent review mentioned the enhanced perceptual abilities of the individuals with this disorder and argues the possibility that these might contribute the survival of the individuals in a Paleolithic context (Spikins et al., 2016). From a cognitive science perspective, on the other hand, here the author presents experimental evidence indicating the gifted cognitive capability of such individuals and suggests the historical presence of a group of professional individuals who had relied upon the capability to survive, being incorporated into human societies. That concerns with the capability of aesthetic judgment of music.

As noted by Temple Grandin (Grandin, 1996; Grandin and Cook, 2004), a strong appreciation for music is commonly observed in children with ASD. Kanner (1943) has already reported extraordinary musical memory in 6 of 11 individuals who were diagnosed as this disorder. The first systematic attempt to identify superior performance on a musical task found that reproduction of atonal melodies was superior in children with ASD, as compared IQ-matched, typically developing (TD) children (Applebaum et al., 1979). Since that study, increased sensitivity to musical pitch and timbre (Heaton et al., 1998), including absolute pitch (AP), has been documented frequently in children with ASD (see Heaton, 2009 for review).

So far, such atypical musical skills have been interpreted in terms of enhanced perceptual processing of lower-order structure of music that is characteristic of this disorder on the assumption that the superiority should not extend to the cognitive sense of music in children with ASD (Mottoron et al., 2000, 2006). However, recently, neurological evidence against the assumption has been presented (Gebauer et al., 2014). In that study, neural correlates of emotional response to music were compared between adults with ASD and neurotypical controls, using functional magnetic resonance imaging. The results revealed that both groups engaged similar neural networks during processing of emotional music. However, in the ASD group, increased activity in response to happy compared to sad music was observed in dorsolateral prefrontal regions as well as in the rolandic operculum/insula, indicating enhanced cognitive processing and physiological arousal in response to emotional musical stimuli in this group. These findings suggest the possibility that an aesthetic emotional response *per se* somehow occurs atypically upon listening to music in individuals with ASD, among whom there are known to be some remarkably musically skillful individuals. This is the issue pursued in the current study.

As a first step, here, the author has compared the aesthetic judgments of consonant/dissonant melody between children with ASD and TD children because this is the issue that has been most intensively as to spontaneous preference for some specific patterns of music. In the Western world as well as in Japan, even newborns are known to show preference for a Mozart's minuet, most of which consists of consonant intervals, over a modified version of it that mostly consists of dissonant intervals (Trainor and Heinmiller, 1998; Masataka, 2006). Thus, the identical testing was undertaken first here with children with ASD and TD children.

While consonant intervals are most often used among popular music around the world, it is also true that some pieces of Western classical music are highly dissonant (Boulez, 1971; Masataka, 2003). Nevertheless, they are appreciated as a source of pleasure. Darwin (1871) already mentioned this fact and was cautious about the connection between musical consonance and hedonicity. For instance, sad and dissonant music like Schoenberg's piano pieces and Adagio by Alibinoni certainly evoke some aesthetic emotions in individuals exposed to them (Boulez, 1971; Thompson et al., 2001; Masataka and Perlovsky, 2013). However, such a pleasurable experience may require a sensitivity that needs to be cultivated as it develops in a particular music culture, unlike preferences for a Mozart's minuet over its modified version with more dissonant intervals. Given this possibility, the degree of the preference for these music pieces in a child, if it is measurable, would reveal the degree to which the musical intelligence defined by the theory of multiple intelligences (Gardner, 2011) has developed in the child. Along such reasoning, in the second experiment of the current study, preferences for such works as well as for typical consonant music such as Mozart's and Bach's pieces were investigated in the two groups of participants.

In order to confirm the major findings of previous studies about unusual musical sensitivity of children with ASD, subsidiary data were collected after the completion of the above experiments from all the participants about AP, the ability to identify the frequency or musical name of a specific tone, or to identify a tone without comparing it with any objective reference tone (Masataka, 2011). The parents of the participants were also interviewed to ask about instances of extraordinary musical memory in their children, which has been reported to occur often in association with ASD (Kanner, 1943).

## MATERIALS AND METHODS

This investigation was conducted according to the principles expressed in the Declaration of Helsinki. All experimental protocols were consistent with the Guide for Experimentation

with Humans, and were approved by the Institutional Ethics Committee, of the Primate Research Institute, Kyoto University (#2011-150). The authors obtained written informed consent from parents of all participants involved in the study.

## Participants

A group of 19 male children with ASD aged 4–7 years ($M = 5.8$; $SD = 1.2$) and 28 TD male children aged 4–7 years ($M = 5.6$; $SD = 1.4$) were studied in the current study. All participants were musically untrained. There was no significant difference between the mean age of the two participant groups [$t(45) = 1.10$, $p = 0.28$]. All participants were Japanese, right-handed, naïve as to the purpose of this study, and auditorily normal.

Nineteen children with ASD were recruited for the current study. Based on direct clinical observation of each child by an independent child psychiatrist, a diagnosis of autism was made according to ICD-10 (World Health Organization, 1994) as well as DSM-IV (American Psychiatric Association, 1994). On the basis of such criteria, each participant in the group of children with ASD was diagnosed as either F84.0, F84.9, or F84.8. Moreover, such diagnoses were also confirmed by the Autism Diagnostic Interview-Revised (ADI-R), an extensive, semi-structured parental interview (Lord et al., 1994) that was conducted by an independent psychiatrist. The ADI-R provides information about the presence of verbal language skills, defined as daily, functional and comprehensive use of spontaneous phrases of at least three words and occasionally a verb. All of the participant ASD children were found to express verbal language. All of the TD children were recruited via the board of education in a small city in Japan. All participants attended normal classes characteristic of their chronological age level. None of the participants included in the groups of TD children met any diagnostic criterion for autism or any other pervasive developmental disorder.

## Procedure

Throughout the current study, two experimental trials were executed to investigate music preference of each participant (referred to as "Experiment 1" and "Experiment 2" below), using the same experimental protocol but different materials as stimuli. None of the participants had been exposed to any of the materials prior to the current study.

The materials used in Experiment 1 were the original, harmonic version of a Mozart's simple minuet (C major K.#1f) and its modified, inharmonic version. They were essentially the same as used previously (Trainor and Heinmiller, 1998; Masataka, 2006; Masataka and Perlovsky, 2012, 2013). Both the original and the modified versions were digitally generated and created by piano timbre. They were made up of 60 intervals. In the original, harmonic version, only three of them were dissonant, and all three were tritons (6-semitone intervals). In the modified, inharmonic version, all Gs were changed to F#s and all Ds to C#s. This had the effect of creating 21 additional dissonant intervals, including a total of 12 of the two most-dissonant intervals, i.e., seven tritons and five minor ninths (13 semitones). In the present stimuli, the upper voice and the lower voice were separated by more than an octave in each interval. The tempo was identical across the two versions.

The materials used in Experiment 2 were four musical pieces, namely, Mozart's piano sonata K.448, Bach's toccata in G major BWV 916, Schoenberg's Klavierstuek op.33a, and Adagio in G minor for organ and strings by Albinoni. While listening to these musical pieces of Mozart and Bach has been reported to enhance spatial tasks, referred to as Mozart effect (Cooper, 2001), the musical pieces of Schoenberg and Albinoni used here are well known as representative of the most popular pieces of highly dissonant western classical music.

As the protocol, the current study adapted the child contingent head-turn preference procedure to allow musical preferences to be measured by the participant pressing a key in order to hear a given extract of music (from a choice of either of two keys/musical pieces in Experiment 1 and from a choice of four possible keys/musical pieces in Experiment 2). The data thus show discrimination among a number of musical styles, and preferences can also be measured. Participants were tested individually, sitting in a quiet room under daylight conditions in front of a table. There was a toy keyboard equipped with eight keys on the table. When a participant entered the room and was seated, an experimenter who was sitting apart from him instructed him that some of the keys would start flashing and that he could hear something by pressing either of them. The keyboard was connected to a portable computer with specially written software, enabling the presentation of each of the prepared stimuli to be linked to each of the flashing keys and to be heard only when the key is pressed. Ten seconds after the instruction, two of the keys started flashing and preliminary trials began. When the participant pressed either of them, a popular Japanese playsong was played. It continued to be played as long as he kept pressing. However, nothing was played when he pressed others keys or more than one key simultaneously. By conducting such trials for a maximum of a 5-min period, each participant learned to perform what was necessary to listen to prepared music stimuli. After the experimenter had confirmed the learning, Experiment 1 was started after a 1-min break. It was 5-min long, and was followed by Experiment 2 of 10-min length after another 1-min break. Thereafter, the participant was instructed to leave the testing chamber. During each break, none of the keys flashed.

In Experiment 1, only two keys of the eight were to flash lights, while four of them were to flash in Experiment 2. The relation of each of the keys and the type of stimulus used was fixed in each experimental trial involving a given participant, randomly determined across the participants, and counterbalanced. The keys were kept flashing during Experiments 1 and 2. During that period, he was allowed to press any key. As long as he kept pressing either of the flashing keys, he was exposed to the music assigned to that key (sound pressure level: 60 dB). However, the music was never played for duration longer than 1 min. If he kept pressing the key longer than 1 min, the computer ceased to present the stimulus. Key presses were recorded in msec, and preference for each music stimulus was calculated as a proportion of the total time allocated to the stimulus.

In testing of AP ability (referred to as Experiment 3), the participants heard 36 pure sine wave tones, presented in pseudorandomized order, which ranged from A3 (tuning: A4 = 440 Hz) to A5, with each tone being presented once. Each tone of the AP test had a duration of 1 s, with a 6-s interstimulus interval. The participants had to answer the tonal label after hearing the accordant tone. Prior to the testing, each participant had been confirmed to have acquired sufficient verbal knowledge about the label to answer. The whole test unit and its components were created with Adobe Audition 1.5. The accuracy was evaluated by counting correct responses. The participants were not asked to identify the adjacent octaves of the presented tones because for AP, identifying the correct chroma is a most notable prerequisite.

In interviews with the parents of the participants, episodes of two children with ASD documented by Kanner (1943) were mentioned; Charles N. was able to distinguish among 18 symphonies before he was 2 years old, and when his mother played his favorite records, he could answer the name of the symphony heard correctly. Another boy, John F., was similarly talented at recognizing melodies. If his father whistled a tune, he often identified it, for instance, as Mendelssohn's violin concerto. In addition, he was able to recite many prayers and nursery rhymes from memory. After such episodes were mentioned to them, the parents in the present study were asked whether they had experienced similar instances in their children.

## RESULTS

The overall results of Experiment 1 are summarized in **Figure 1**, which shows the overall mean duration of listening to the original, harmonic version of the Mozart's minuet and to its modified, inharmonic version with many dissonant intervals in the group of TD children and that of ASD children. When the collected data were analyzed using a 2 (ASD/TD, PARTICIPANT) × 2 (original version versus modified version, STIMULUS) analysis of variance (ANOVA), one of the two main effects (STIMULUS) was statistically significant [$F(1,45) = 49.53$, $p = 0.0000$, $\eta_p^2 = 0.299$]. The other main effect (PARTICIPANT) was not significant [$F(1,45) = 0.84$, $p = 0.46$, $\eta_p^2 = 0.033$]. However, the interaction between PARTICIPANT and STIMULUS was also significant [$F(1,45) = 4.25$, $p = 0.028$, $\eta_p^2 = 0.35$].

Subsequent analyses of simple main effects (Bonferroni correction) revealed that the mean listening time was shorter in response to the inharmonic version than to the harmonic version in both the children with ASD and TD children ($p < 0.001$). Moreover, the children with ASD listened to the harmonic version longer than the TD children ($p = 0.03$), who listened to the inharmonic version longer than the children with ASD ($p = 0.04$).

The overall results of Experiment 2 are summarized in **Figure 2**, which shows the overall mean duration of listening to the pieces of music from the four classic works in the group of TD children and that of ASD children. When the data were analyzed using another ANOVA, both of the two main effects were statistically significant [$F(1,45) = 48.33$, $p = 0.0000$, $\eta_p^2 = 0.287$ for STIMULUS and $F(1,45) = 9.12$, $p = 0.009$, $\eta_p^2 = 0.133$ for PARTICIPANT]. The interaction between PARTICIPANT and STIMULUS was also significant [$F(3,135) = 4.25$, $p = 0.012$, $\eta_p^2 = 0.30$].

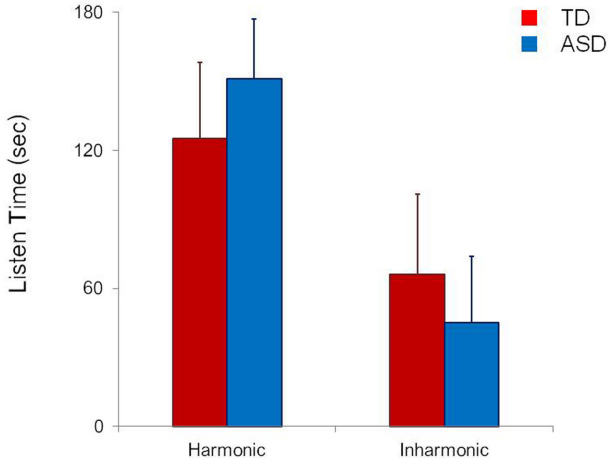



**FIGURE 1 |** Mean preference (error bars: SDs) for an original simple harmonic minuet of Mozart (Harmonic) and its modified inharmonic version containing many dissonant intervals (Inharmonic) in children with autism spectrum disorder (ASD) and in typically developing (TD) children.

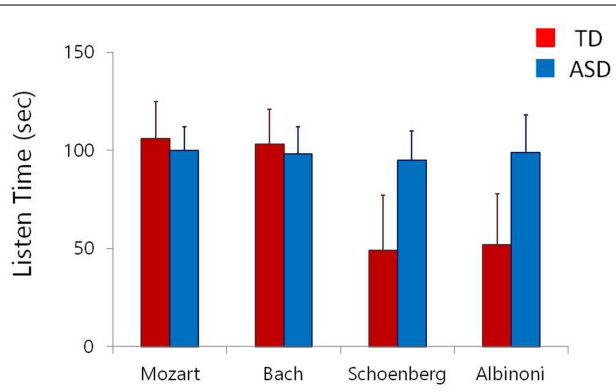



**FIGURE 2 |** Mean preference (error bars: SDs) for musical pieces of Mozart, Bach, Schoenberg and Albinoni in children with autism spectrum disorder (ASD) and in typically developing (TD) children.

Pair-wise comparisons of the mean listening time across the four different music stimuli in the TD children and the children with ASD revealed that the TD children listened longer to both Mozart's piece and Bach's piece than to either Schoenberg's piece or Albinoni's piece ($ps < 0.01$). Besides these, no such differences were found with regard to any pair among the four stimuli ($ps > 0.10$). The mean listening time of the children with ASD to any of the four stimuli did not significantly differ from that to any of the remaining three stimuli ($ps > 0.10$). Overall the children with ASD listened to Schoenberg's piece and Albinoni's piece longer than the TD children ($ps < 0.01$) whereas the

mean listening time did not differ significantly between the two participant groups with regard to Mozart's piece or Bach's piece ($ps > 0.10$).

The overall results of the subsequent AP testing (Experiment 3) are presented in **Figure 3**. While the accuracy scores of all of the TD children remained within the level of chance, those of 15 of the 19 children with ASD were found to exceed the level of chance. During the interviews with the parents of the participants, some instance of extraordinary musical memory in their children was reported by six parents of the children with ASD, but there was no such report by any parent of the TD children.

## DISCUSSION

Regarding TD children, the results of Experiment 1 of the current study are consistent with those reported previously (Trainor and Heinmiller, 1998; Masataka, 2006) and these participants exhibited a preference for the original, harmonic version of a Mozart's minuet over its modified version containing many dissonant intervals. The results of this experiment also showed the fact that such a preference was confirmed in children with ASD. Moreover, the extent of the preference was more robust in children with ASD than in TD children, indicating that children with ASD are more sensitive than TD children are with regard to the aesthetic judgment of consonance and dissonance *per se* in music.

The results of Experiment 2 are seemingly to contradictory to those of Experiment 1. Namely, children with ASD listened to the music stimuli indiscriminately whether the music consisted of many dissonant intervals or not whereas TD children listened to the pieces of music consisting of many consonant intervals longer than to those consisting of dissonant intervals, to which the TD children actually avoided listening. Due to such avoidance, the total duration of listening to the stimuli was longer in children with ASD than in TD children. However, it should be noted that the pieces of music containing many dissonant intervals that were used as the stimuli in Experiment 2 were from famous composers of Western classical music. These pieces contained such dissonant intervals as a consequence of the composers' attempt to aesthetically express the emotion of sadness, whereas in Experiment 1, Mozart's original minuet was merely modified using a means to create its dissonant version.

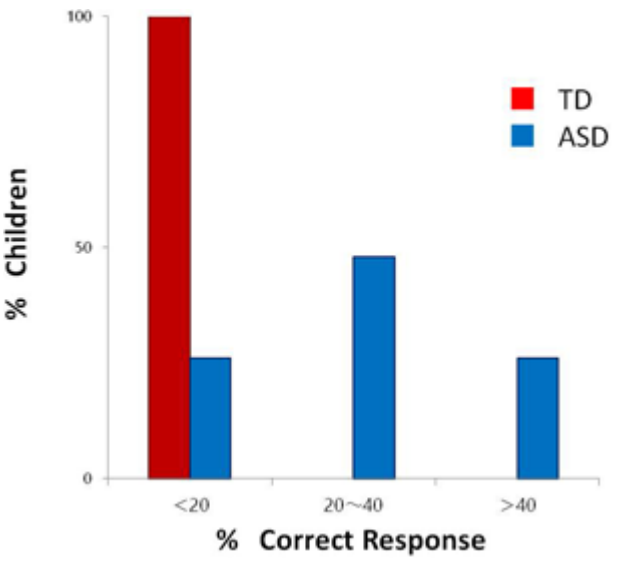



**FIGURE 3 |** Accuracy scores in absolute pitch testing in children with autism spectrum disorder (ASD) and in typically developing (TD) children.

This difference differentially influenced the responses to the stimuli in children with ASD, but not in TD children. This fact appears to deserve further analysis. Possibly, it may point to a question about the validity of the general assumption that the relative proportion between consonance and dissonance in music corresponds to a hedonic division between positive and negative that, apparently, in turn corresponds to the distinction between major and minor in music. The present results suggest that children with ASD possess enhanced capability of aesthetic judgment and appreciated the presented pieces of classical music even though they contained many dissonant intervals. This, in turn, indicates that it would be difficult to explain the more robust response to consonance/dissonance difference in children with ASD than in TD children found in Experiment 1 merely in terms of an enhanced perceptual processing of lower-order structure of music that is associated with ASD.

Among a number of acoustical dimensions constituting music, that of consonance/dissonance is the developmentally earliest that can be judged. During early childhood, neurotypical individuals are likely to listen to harmonic music preferably to music with many dissonant intervals even if those intervals in the latter music were a consequence of an artistic attempt to express the emotion of sadness aesthetically (Winner, 2006). Appreciation of sad music develops later. While this was indeed experimentally confirmed in Experiment 2 of the current study, the present results also raise the possibility that children with ASD who are aged-matched to TD children possess a more developed capability of aesthetic judgment of music even though musically untrained.

Along with the conceptualization of the theory of multiple intelligences (Gardner, 2011), the cognitive characteristics revealed in children with ASD can be explained as greater than typical strength of musical intelligence in themselves although the children experience difficulty with interpersonal communication (Lord et al., 1994; World Health Organization, 1994). While the pattern of operation of an individual person's mind can be categorized according to the domain toward which that individual is more oriented, individuals with ASD, overall, do not rely upon their social relationships but rather are predisposed to process perceived non-social objects in more depth (Masataka, 2017). For instance, these individuals "exhibit enhanced discrimination between auditory stimuli, more accurate local target detection of auditory stimuli, and diminished global interference with auditory processing." These abilities, referred to as "naturalist intelligence" by Gardner (2011) are highly adaptive for living in nature, and in particular,

would seem to have conferred an evolutionary advantage upon individuals during prehistoric times, as well as upon individuals in later times who would prefer to live outside any community (Armstrong, 2017).

The enhanced capability of aesthetic judgment of music found here was associated with some atypical musical skills characteristic of ASD such as AP and extraordinary musical memory, as revealed by the subsidiary data collected after the completion of Experiments 1 and 2. These skills are known to have been possessed by an outstanding artist in the 16th century, Wolfgang Mozart, who could name the note on hearing a bell toll, or a clock or even a pocket watch strike, and could memorize a 10-min-long melody that he heard only once (Deutsch, 2006). It should also be noted that his family kept continually touring throughout Europe from the time he was 5 years until 21 years of age, which contributed to establishing the young composer's reputation as a musical prodigy. Such a lifestyle was indeed in accord with the lifestyles of such individuals who were documented in historical records in Europe as professional musical performers (Hartung, 2003).

The documentation of such individuals in southern Europe, the region between the Alps and the Pyrenees, appeared as early as the 10th century, with them being referred to as "Spielmann (plural: Spielleute)" in German and as "wandering minstrels" in English. Spielleute had no settled abode, and instead used to roam about from place to place over a broad range of regions across a number of communities. In the 13th century, such individuals came to be distributed over all regions of Europe. While all other individuals who were integrated into feudal society were occupied with predetermined, inherited work, Spielleute were the first who had chosen this occupation of their own will. Born in urban regions in the society, most of them felt difficulty in staying there and decided to live as outcasts. As a result, they were treated as being dishonorable by society members. They were talented in a variety of rudimentary forms of contemporary art performance, performing arts including music and dancing without experiencing any form of professional training. In particular, they were well known to perform mimicking of a variety of sounds heard in the landscape (such as bird songs) while playing stringed and/or woodwind instruments simultaneously (Hartung, 1982). When one of their performances was successful, the melody was memorized no matter how long it was without recording it using musical notation (actually, they were unable to read music) and it became part of their repertoire of routine performance. Such a seemingly mysterious endowment seems also to have been possessed by some famous modern artists who contributed to the subsequent development of classical music, such as Mozart, Bach and Beethoven (Deutsch, 2006), and also by the children with ASD in the current study.

Spielleute were allowed to travel around as they liked by the churches and the lords of communities, their daily income totally depending upon their performances. They were apparently regarded as exceptional individuals (outsiders). The official position of the Catholic church was that they lived in a state of sin (Stuart, 1999). Nevertheless they used to play an important role, i.e., the role of 'negotiator' once conflict happened between regional communities within feudal society because they were familiar with members of both of the disputing communities. That indicates that despite being socially marginalized, such individuals were socially functional. While "Spielleute" or "wandering minstrels" were neurdivergent individuals who came to be no longer recorded after the 18th century as Europe became industrialized, the fact should be noted that similar outcast groups of individuals have been commonly documented among a variety of cultures and that all of them were talented in musical performance without exception (Hartung, 2003). That again strongly suggests the possibility that they possessed an enhanced aesthetic sense of music such as the giftedness that was revealed here in the performance of children with ASD.

As clinical implications of the findings, one can also mention the possibility that this early onset and/or potentially predispositional demonstration of heightened aesthetic judgment of music might eventually lead to new job-related opportunities for individuals with ASD in uniquely useful and professionally applicable settings in new fields. No doubt, this is an issue that should be clinically addressed in the next step of research.

## AUTHOR CONTRIBUTIONS

NM designed the study, collected and analyzed the data and drafted the manuscript.

## FUNDING

The study was supported by Japan Society for the Promotion of Science London and the grant-in-aid number is #25285201.

## ACKNOWLEDGMENTS

The author is greatly indebted to Thomas Armstrong for his number of thoughtful comments on the earlier version of the manuscript. The author is also grateful to Hiroki Koda for assistance in conducting the experiments and to Elizabeth Nakajima for proofreading the English of the manuscript.

## REFERENCES

American Psychiatric Association (1994). *Diagnostic and Statistical Manual ofMental Disorders*. Washington DC: American Psychiatric Association, 1–609.

Applebaum, E., Egel, A. L., Koegel, R. L., and Imboff, B. (1979). Measuring musical abilities of autistic children. *J. Autism Dev. Disord.* 9, 279–285. doi: 10.1007/BF01531742

Armstrong, T. (2012). *Neurodiversity in the Classroom: Strength-based Strategies to Help Students with Special Needs Succeed in School and Life*. Alexandria, VA: ASCD, 1–183.

Armstrong, T. (2017). The healing balm of nature: understanding and supporting the naturalist intelligence in individuals diagnosed with ASD. *Phys. Life Rev.* 20, 109–111. doi: 10.1016/j.plrev.2017.01.012

Boulez, P. (1971). *Boulez on Music Today*. London: Faber and Faber, 1–156.

Cooper, J. S. (2001). The mozart effect. *J. R. Soc. Med.* 94, 170–172.
Darwin, C. R. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray, 1–265.
Deutsch, D. (2006). The enigma of absolute pitch. *Acoust. Today* 2, 11–19. doi: 10.1121/1.2961141
Gardner, H. (2011). *Frame of Mind: The Theory of Multiple Intelligences*. New York, NY: Basic Books, 1–387.
Gebauer, L., Skewes, J., Westphael, G., Heaton, P., and Vuust, P. (2014). Intact brain processing of musical emotions in autism spectrum disorder, but more cognitive load and arousal in happy vs sad music. *Front. Neursci.* 8:192. doi: 10.3389/fnins.2014.00192
Grandin, T. (1996). *Thinking in Pictures*. New York, NY: Vintage, 1–177.
Grandin, T., and Cook, K. (2004). *Developing Talents: Careers for Individuals with Asperger Syndrome and High-functioning Autism*. Lenexa, KS: Autism Asperger Publishing, 1–185.
Hartung, W. (1982). *Die Spielleute: Eine Randgruppe in der Gesellschaft des Mittelalters*. Wiesbaden: Franz Steiner Verlag, 1–108.
Hartung, W. (2003). *Die Spielleute im Mittelalter: Gaukler, Dichter, Musikanten*. Dusseldorf: Artemis & Winkler, 1–405.
Heaton, P. (2009). Assessing musical skills in autistic children who are not savants. *Philos. Trans. R. Soc. B* 364, 1443–1447. doi: 10.1098/rstb.2008.0327
Heaton, P., Hermelin, B., and Pring, L. (1998). Autism and pitch processing: a precursor for savant musical ability? *Music Percept.* 15, 291–305. doi: 10.2307/40285769
Kanner, L. (1943). Autistic disturbances in affective contact. *Nerv. Child* 2, 217–250.
Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord.* 24, 659–685. doi: 10.1007/BF02172145
Masataka, N. (2003). *The Onset of Language*. Cambridge: Cambridge University Press, 1–281. doi: 10.1017/CBO9780511489754
Masataka, N. (2006). Preference for consonance over dissonance by hearing newborns of deaf parents and of hearing parents. *Dev. Sci.* 9, 46–50. doi: 10.1111/j.1467-7687.2005.00462.x
Masataka, N. (2011). Enhancement of speech-relevant auditory acuity in absolute picth possessors. *Front. Psychol.* 2:101. doi: 10.3389/fpsyg.2011.00101
Masataka, N. (2017). Implications of the idea of neurodiversity for understanding the origins of developmental disorders. *Phys. Life Rev.* 20, 85–108. doi: 10.1016/j.plrev.2016.11.002
Masataka, N., and Perlovsky, L. (2012). The efficacy of musical emotions evoked by Mozart's music for the reconciliation of cognitive dissonance. *Sci. Rep.* 2, 307. doi: 10.1038/srep00307
Masataka, N., and Perlovsky, L. (2013). Cogntive interefernce can be mitigated by consonant music and facilitated by disoonant music. *Sci. Rep.* 3:2028. doi: 10.1038/srep02028
Mottoron, L., Dawson, M., Soulieres, L., Hubert, B., and Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *J. Autism Dev. Disord.* 2, 1–17. doi: 10.1007/s10803-005-0040-7
Mottoron, L., Peretz, I., and Menard, E. (2000). Local and global processing of music in high-functioning persons with autism: beyond central coherence? *J. Child Psychol. Psychiatry* 41, 1057–1065. doi: 10.1111/1469-7610.00693
Silberman, S. (2015). *Neurotribes: The Legacy of Autism and the Future of Neurodiversity*. New York, NY: Avery, 1–534.
Spikins, P., Wright, B., and Hodgson, D. (2016). Are there alternative adaptive strategies to human pro-sociality? The role of collaborative morality in the emergence of personality variation and autistic traits. *Time Mind* 9, 289–313. doi: 10.1080/1751696X.2016.1244949
Stuart, K. (1999). *Defiled Trades Social Outcasts: Honor and Ritual Pollution in Early Modern Germany*. Cambridge: Cambridge University Press, 1–286.
Thompson, W. F., Schellenberg, E. G., and Husain, G. (2001). Arousal mood and the Mozart effect. *Psychol. Sci.* 12, 248–251. doi: 10.1111/1467-9280.00345
Trainor, L., and Heinmiller, B. M. (1998). The development of evaluative response to music: Infants prefer to listen to consonance over dissonance. *Infant Behav. Dev.* 21, 77–88. doi: 10.1016/S0163-6383(98)90055-8
Winner, E. (2006). "Development in the arts: drawing and music," in *Handbook of Child Psychology*, Vol. 2, ed. R. Damon (New York, NY: Wiley), 859–904.
World Health Organization (1994). *The Composite International Diagnostic Interview, Version 1.1*. Geneva: World Health Organization, 1–632.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Masataka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

PERSPECTIVE
published: 22 September 2017
doi: 10.3389/fpsyg.2017.01581

# What Vowels Can Tell Us about the Evolution of Music

*Gertraud Fenk-Oczlon**

*Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria*

Whether music and language evolved independently of each other or whether both evolved from a common precursor remains a hotly debated topic. We here emphasize the role of vowels in the language-music relationship, arguing for a shared heritage of music and speech. Vowels play a decisive role in generating the sound or sonority of syllables, the main vehicles for transporting prosodic information in speech and singing. Timbre is, beyond question, the primary parameter that allows us to discriminate between different vowels, but vowels also have intrinsic pitch, intensity, and duration. There are striking correspondences between the number of vowels and the number of pitches in musical scales across cultures: an upper limit of roughly 12 elements, a lower limit of 2, and a frequency peak at 5–7 elements. Moreover, there is evidence for correspondences between vowels and scales even in specific cultures, e.g., cultures with three vowels tend to have tritonic scales. We report a match between vowel pitch and musical pitch in meaningless syllables of Alpine yodelers, and highlight the relevance of vocal timbre in the music of many non-Western cultures, in which vocal timbre/vowel timbre and musical melody are often intertwined. Studies showing the pivotal role of vowels and their musical qualities in the ontogeny of language and in infant directed speech, will be used as further arguments supporting the hypothesis that music and speech evolved from a common prosodic precursor, where the vowels exhibited both pitch and timbre variations.

**Keywords: vowels, timbre, musical scale, non-Western cultures, ethnomusicology, musical protolanguage, evolution of music, linguistic typology**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky,*
*Braavo! Enterprises, United States*

***Reviewed by:***
*John G. Neuhoff,*
*College of Wooster, United States*
*Bahia Guellai,*
*University of Paris Ouest Nanterre, France*

***Correspondence:**
*Gertraud Fenk-Oczlon*
*gertraud.fenk@aau.at*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *25 May 2017*
***Accepted:*** *29 August 2017*
***Published:*** *22 September 2017*

***Citation:***
*Fenk-Oczlon G (2017) What Vowels Can Tell Us about the Evolution of Music. Front. Psychol. 8:1581. doi: 10.3389/fpsyg.2017.01581*

## INTRODUCTION

The evolution of music, the evolution of language and possible common evolutionary pathways of these achievements remain a matter of debate. Is music only a non-adaptive by-product of language as suggested by Pinker (1997) or did "previously developed musical powers" (Darwin, 1871, p. 12) precede the oratory? An intriguing hypothesis is that music and language share a common origin in form of a musical protolanguage in which song-like strings had holistic meanings (e.g., Jespersen, 1894; Mithen, 2005; Fitch, 2006).

Here, we emphasize the role of vowels in the music-language relationship[1], and present arguments that reinforce the hypothesis of a prosodic protolanguage or 'protomusic' (Fitch, 2010). Vowels play a decisive role in generating the sound or sonority of syllables, the main vehicles for

[1]There are, of course, many other areas where music and language show similarities. For example, striking rhythmic relationships between music and speech (Patel and Daniele, 2003; Neuhoff and Lidji, 2014) or correspondences in the length of 'intonation units' (Fenk-Oczlon and Fenk, 2009b; Lehmann and Goldhahn, 2016).

transporting prosodic information in speech, and singing. In tone languages, which represent more than half the world's languages, vowels carry the pitch modulations that convey grammatical and lexical information. The tight bond between vowels and pitch is supported by experimental findings suggesting strong interactions in the processing of vowels and melody, but not between consonants and musical information: "Vowels sing but consonants speak" (Kolinsky et al., 2009, p. 1). Likewise, a mismatch negativity (MMN) study by Lidji et al. (2010) revealed a close processing relationship between vowels and pitch even at a pre-attentive level.

This perspective paper starts with a brief theoretical analysis of the sound systems of language and music, highlights general parallels in sound inventories, and reports coincidences between vowel systems and musical scales even in specific cultures. Moreover, we demonstrate a close match between vowel pitch and musical pitch in nonsense syllables of Alpine yodelers and in the yodeling of African Pygmies. We will discuss these findings in the context of other ethnomusicological findings, frequently used (Wiora, 1962; Nikolsky, 2015) to shed light on prehistoric music or on the origin of music. Studies showing the pivotal role of vowels and their musical qualities in the ontogeny of language and in infant directed speech will be used as further arguments for the musical protolanguage hypothesis.

## SOUND SYSTEMS IN MUSIC AND LANGUAGE: PITCH AND TIMBRE

Jackendoff (2009, p. 198) emphasizes the differences in the sound system of language and music: "In phonological structure, the repertoire of speech sounds forms a structured space of timbres /. . ./. In music, by contrast, the notes are distinguished by the way they form a structured space of pitches. . ."

But the distinction between timbre contrasts in speech on the one hand and pitch contrasts in music on the other is not always as clear-cut. Especially in some non-Western music, vocal timbre and musical melody are often intertwined. Walker (1997) provides evidence that Canadian Inuit, West Coast Native people, and Australian Aborigines do not distinguish between melody and vocal timbre when discussing their music. In the language of the Inuit, for instance, there does not even exist a word that differentiates between language and music: *nipi* means music and sound of the spoken voice (Nattiez, 1990). And microtonal pitch intervallic movements that could be related to vocal timbre changes (Walker, 1997) are characteristic for Australian Aboriginal music. Vocal timbre and timbral variations are also the main idea behind khasmatonal music whose tonal organization can be characterized as "half-spoken/half-sung with intense timbral/pitch modifications" (Nikolsky, 2015, p. 13). In addition to vowel timbre, which seems to be pivotal in timbre-driven music, a person's voice timbre (Sheikin, 2002) is an important and often exaggerated element in the music of many Siberian ethnicities.

### Vowels: Timbre, Intrinsic Pitch, Intensity, Duration

Timbre is, beyond question, the primary parameter that allows discriminating between different vowels and to some degree also between different consonants (e.g., nasals, liquids). The human voice, as well as the sounds of most musical instruments, is "made up of many nearly harmonically related frequency components, or partials" (Pierce, 1999, p. 8). The timbre of sounds is determined by the distribution of the sound energy among partials (overtones) of different frequencies.[2] Vowels differ from each other by specific peaks or formants in their sound spectra, whereas the formants F1 and F2 are most relevant for their identification (Peterson and Barney, 1952). The formants correspond to the resonances of the vocal tract or oral cavity[3]; the main articulatory parameters responsible for vowel timbre are tongue height, front-to back position of the tongue, and lip rounding.

But vowels not only differ in their timbre but also in their intrinsic pitch, intensity, and duration. It is generally postulated that open (low) vowels have a higher intrinsic duration and a higher intrinsic intensity than close (high) vowels (for more details see Möbius, 2003). Concerning vowel intrinsic pitch, it is known since Meyer (1896) that high vowels such as [i] have a higher intrinsic fundamental frequency (IF0) than low vowels such as [a]. While the mechanism which determines IF0 is still a subject of debate, there seems to be general agreement that vowel pitch depends primarily on the frequency of the second formant F2 (Marks, 1978; Traunmüller, 1986). Vowels with a high F2 (e.g., [i], [y]) have a higher intrinsic pitch than vowels with a low F2 (e.g., [a], [o]). The close association between F2 or spectral energy allocation and vowel intrinsic pitch indicates that timbre and intrinsic pitch of vowels are closely interrelated and cannot be separated.

### Vowel Timbre and Musical Melody

In Fenk-Oczlon and Fenk (2009b) we hypothesized that in songs containing strings of meaningless syllables the vowels might be connected to melodic direction in close correspondence to their timbre or intrinsic pitch. We tested this assumption based on all monophonic Alpine yodelers ($n = 15$) in Pommer's (1893) collection. The test revealed a surprisingly uniform pattern: the melody descended in 118 out of 121 [i]→[o] successions and ascended in 132 out of 133 [o]→[i] successions. A similar assumption was tested in Austrian traditional songs, which include successions of nonsense syllables. In 24 out of 26 songs in the Dawidowicz (1980) collection we found the expected coincidence between the vowel [i] and the highest pitch in melody (Fenk-Oczlon and Fenk, 2009a).

A strong relationship between vowel timbre and musical pitch in meaningless syllables is also reported in the yodeling of

[2]In fact, timbre does not only depend on spectral aspects of sound but also on time-varying aspects, i.e., the amplitude envelope (e.g., Caclin et al., 2005).

[3]For a detailed description of the human vocal tract, its relationship to vowel formant frequencies and potential relevance in the evolution of human language (see, e.g., Fitch, 2010), and a recent discussion of this interesting topic by Fitch et al. (2017) and Lieberman (2017).

African Pygmies (Fürniss, 1991; Demolin, 2013): front vowels are associated with high pitch and back vowels with low pitch.

Vowel timbre, moreover, plays a key role in transforming spoken information into whistled languages (Meyer, 2008), in 'talking khomus' using the jew's harp to transmit verbal information in Yakut traditional music (Alexeyev and Spiridon, 2004), as well as in many mnemonic systems for transmitting or representing musical melodies (Hughes, 2000).

## SOUND SYSTEMS IN LANGUAGE AND MUSIC: PARALLELS IN THE INVENTORY SIZE

Are there parallels in the sound inventory size of language and music? Authors looking for parallels in the sound inventories of language and music often compared the whole phonemic inventory to musical pitches per octave and found that the number of phonemes across languages varies to a much greater extent ["from 11 in Polynesian to 141 in the languages of the Bushmen" (Besson and Schön, 2001, p. 235)] than the number of pitches per octave. And Rakowski (1999) argues that the number of phonemes in languages is much higher than the number of musical intervals, which roughly corresponds with Miller's magical number seven. An alternative approach was provided in Fenk-Oczlon and Fenk (2009b), comparing only the vowel inventories with the number of intervals in a musical scale, instead of the whole phonemic inventory.

Concerning the vowel inventories across languages there is unanimous consensus that five-vowel systems are the most frequent ones, followed by six and seven vowel-systems (Crothers, 1978; Schwartz et al., 1997; Ladefoged, 2005; Maddieson, 2005). For instance, Maddieson's (2005) sample in the World Atlas of Language Structures (WALS) comprises 563 languages; the smallest vowel quality inventory is 2 and the largest is 14. Most of the languages have five vowels, followed by six and seven. Four languages have two contrasting vowel qualities, only one language (German) has 14 basic vowels, and only two languages (British English; Bété) make use of 13 vowels. Note, however, that in the respective samples only basic vowel qualities are counted, and variations of basic vowel qualities such as nasalisation, pharyngalisation, length, etc. are not considered.

Unfortunately, statistical databases for scale types across cultures, analogously to those of vowel inventories, do not seem to exist. Musical scales "are classified according to the number of tones used, their range, and their intervals" (Nettl, 1956, p. 46). The simplest scales are diatonic (Nettl, 1956), and Burns (1999) argues that 12 pitches per octave represent a practical limit. The existence of a higher number of tones – e.g., through "intervals that bisect the distance between the Western chromatic intervals" – is, at least as a standard in the culture in question (the Arab-Persian system), a rather controversial question (Burns, 1999, pp 217–218). There is a general agreement that five-tone (pentatonic) scales are the most frequent ones amongst traditional forms of music, and that musical scales across cultures typically have five to seven pitches (Trehub et al., 1999, p. 965). In contrast to language, where six-vowel systems appear to be slightly more frequent than seven-vowel systems, the seven-tone scale (heptatonic scale) seems to be more frequent than the six-tone (hexatonic) scale. Concerning the lower number of six-note scales in comparison to five- or seven-note scales, Gill and Purves (2009) argue that six-note variants of five- or seven-note scales are very frequent e.g., in blues scales, or in melodies using only six out of the seven tones of the heptatonic scale, or using passing tones not included in the pentatonic scale. They assume that six-note scales "are simply not recognized as formally as their five- and seven-note counterparts in Western music theory" (p. 8). In non-Western musical cultures pentatonic scales are, as in Western music, the most frequent ones; but it is interesting to see, that they use, according to Nettl (1956, p. 60), hexatonic scales more frequently than heptatonic scales.

We here state some striking coincidences in the sound inventories of language and music: an upper limit of roughly 12 elements, a lower limit of 2 elements, and a frequency peak at 5–7 elements. It should be noted, that in our comparison between vowels systems and musical sales we only considered the number of elements, and we did not consider the patterning of intervals in a scale or the distribution of vowels in the vowel space.

### Are There Correspondences between Vowels and Scales Even in Specific Cultures?

In Fenk-Oczlon and Fenk (2009a) we speculated that there might be coincidences between vowels and scales even in specific cultures. Consistently, most of the Australian Aboriginal languages have three vowels, only those of the Northern Territory are reported to have more vowels (Butcher and Anderson, 2008). And Lauridsen (1983) reports exactly from those cultures in the Northern Territory the use of a higher number of pitches in song. The most difficult part in testing our assumption is to establish which use of certain musical features is indigenous, and which comes as cultural borrowing. Therefore, looking for coincidences between the number of vowels and number of pitches may be particularly informative in cultures before they had extended contact with other musical traditions. Some more evidence we found so far:

The indigenous cultures of the Americas – such as the Arapaho, Blackfoot, Cheyenne, Comanche, Klamath, Modoc, Navaho, Pawnee – frequently have only four or three vowels (Maddieson, 2005), and tritonic and tetratonic scales are frequently used in their music (Rhodes, 1954; Nettl, 1956). Noteworthy, Nettl (1956, p. 112) reports that in Pueblo music which includes the music of the Hopi, Zuni, and Taos "the scales tend to have more tones than they do elsewhere on the continent." And Hopi, Zuni, and Taos belong exactly to those rather rare indigenous languages of the Americas which have five or six vowels.

Correspondences between three-vowel systems and tritonic scales show in the Inca musical tradition and in the music of Greenland. Pre-Hispanic herranza ritual music of the Andes

is generally tritonic –rather than "incomplete pentatonic" (Holzmann, 1980), and Incan Quechua as well as other Quechuan varieties have just three vowels. Greenlandic like the vast majority of Inuit languages has only three vowels, and tritonic modes are still common in the music of East Greenland (Olsen, 1972). Moreover, tritonic patterns are supposed to have been characteristic to Inuits from Greenland to Siberia.

The correspondences found between vowel inventories and musical scales are, of course, highly tentative. Nevertheless, it would be interesting to investigate whether these correspondences can also be found in other autochtonous cultures.

## VOWELS IN ONTOGENY, PHYLOGENY, AND MOTHERESE

Vowels and their musical qualities play a pivotal role in the ontogeny of language and in infant directed speech. Vowel-like sounds are the first speech sounds that children produce and are already present in the cooing stage appearing at about 6–8 weeks of age. "The first coos that infants make sound like one long vowel" and the infant's learning process to be undertaken in this stage is to "produce a series of different vowel-like sounds strung together but separated by intake of breath" (Hoff, 2009, p. 143). The coos exhibit pitch variations, which according to Halliday (1975, cited in Masataka, 2007) are already connected with pragmatic functions. For instance: requests for objects are associated with a rising pitch, and labeling is associated with falling pitch. Masataka (2007) argues that pitch variations of this sort might be strengthened by caregivers in using infant directed speech or motherese, which can be above all characterized by high pitches and exaggerated pitch contours. Thus, infant directed speech is generally considered having many musical qualities and providing a scaffold for language acquisition (Falk, 2004). Again it is the vowels that are acoustically exaggerated and the more caregivers are exaggerating the vowels, and the larger the vowel space they use, the better the infant's language performance (Liu et al., 2003). The importance of vowel pitch or vowel timbre in early communication is demonstrated by Papousek's (1992) findings (cited in Walker, 1997) where infants' abilities to discriminate pitch variations in the mother's voice is linked to "vowel" pitch, not "musical" pitch.

But vowels or vowel-like sounds may not only play a prominent role in the ontogeny of human language, but also in its early evolution. Notably, recent work by Boë et al. (2017) demonstrated astonishing similarities between human vowels and "vowel like segments" produced by Guinea baboons. Combining acoustical analyses of baboon vocalizations with an anatomical study of the animals' vocal tract, the authors showed that baboons are capable to produce and combine five vowel-like sounds, despite their high larynx, suggesting a proto-vocalic system was already present in the last common ancestor of humans and baboons, at least 25 million years ago (Boë et al., 2017).

## DISCUSSION AND CONCLUSION

We started with a theoretical analysis of the sound systems of language and music and and demonstrated a closer relationship and more parallels between these systems than generally assumed. Vowels show all the core properties of music, i.e., timbre, intrinsic pitch, intensity, and duration. We revealed general correspondences between vowel systems and musical scales across cultures, and presented correspondences even in specific cultures. While the general parallels in the sound inventories may reflect cognitive and physiological constraints of our auditory and articulatory apparatus, the coincidences found between number of vowels and number of tones in specific cultures indicate a tight bond between vowels and music. The match between vowel pitch and musical pitch in meaningless syllables of Alpine yodelers and yodelers of African Pygmies and the relevance of vocal timbre in the music of many non-Western cultures, in which vocal timbre/vowel timbre and musical melody are often intertwined, further support a close relationship between vowels and music. How can these findings be related to the evolution of music?

The evidence we provided for a close relationship between vowels and musical pitch stems from studies of ethnic or non-Western music. And ethnomusicological findings are often assumed to particularly revealing with regards to the origin of music. For example, Nikolsky (2015) uses ethnic music to reconstruct prehistoric music and assumes that the earliest music "was organized not by pitch, but by timbre" (p. 25). This assumption implicates that at this early stage, music and speech must have been together and that it was the vowels which displayed – besides voice timbre – timbre and pitch modulations.

Taking further into account the pivotal role of vowels in ontogeny – phylogeny in some respects and to some degree resembles ontogeny and vowel-like sounds are the first speech sounds that children produce – it is tempting to speculate that the earliest human vocal communication started with vowels or vowel syllables strung together, which were connected by semivowels or glides such as [w], [h], [j] or the glottal stop [ʔ]. The sequences of vowels exhibited pitch and timbre modulations, which were combined for different pragmatic and social functions, and were probably propositionally meaningless (cf. also Fitch, 2010). In a later stage, more 'real' consonants such as obstruents emerged and were combined with vowels into consonant-vowel syllables. This was likely the emergence of articulated speech (Jordania, 2006) and of utterances which could express propositional meanings.

We still encounter vocal music in which mere vowel sounds are connected. For example, Lewis (2013) reports that the songs of the BaYaka Pygmies rarely have words, but many songs are based on vowel-sound melodies or hocketed vowel sounds.

In conclusion, we have demonstrated a close relationship between vowels and music in non-Western cultures which may shed light on the earliest human vocal communication and may strengthen the idea of a musical protolanguage. Although

our findings are preliminary, we are convinced that future research combining ethnomusicological findings with those from linguistic typology will provide further insight into the evolution of music, and in music as such.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## REFERENCES


Alexeyev, I., and Spiridon, S. (2004). The music of the yakut traditional khomus. *J. Int. Jews Harp Soc.* 1, 88–94.

Besson, M., and Schön, D. (2001). Comparison between language and music. *Ann. N. Y. Acad. Sci.* 930, 232–258. doi: 10.1111/j.1749-6632.2001.tb05736.x

Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T., et al. (2017). Evidence of a vocalic proto-system in the baboon (papio papio) suggests pre-hominin speech precursors. *PLOS ONE* 12:e0169321. doi: 10.1371/journal.pone.0169321

Burns, E. M. (1999). "Intervals, scales and tuning," in *The Psychology of Music*, 2nd Edn, ed. Diana Deutsch (San Diego, CA: Academic Press), 271–300.

Butcher, A., and Anderson, V. B. (2008). "The vowels of Australian aboriginal english," in *Proceedings of the 9th Annual Conference of the International Speech Communication Association*, Brisbanen, QLD, 347–350.

Caclin, A., McAdams, S., Smith, B. K., and Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *J. Acoust. Soc. Am.* 118, 471–482. doi: 10.1121/1.1929229

Crothers, J. (1978). "Typology and universals of vowel systems," in *Universals of Human Language*, ed. J. H. Greenberg (Stanford, CA: Stanford University Press), 93–152.

Darwin, C. H. (1871). *The Descent of Man and Selection in Relation to Sex*, 1st Edn. London: John Murray.

Dawidowicz, A. (1980). *Österreichisches Liederbuch. Komm, Sing Mit!* Innsbruck: Musikverlag Helbling.

Demolin, D. (2013). *Pitch and Timbre in the Languages and Music of Central African Pygmies*. Available at: http://media.leidenuniv.nl/legacy/workshop-pitch-and-timbre-in-the-language-music-relation.pdf

Falk, D. (2004). Prelinguistic evolution in early hominins: whence motherese? *Behav. Brain Sci.* 27, 491–541. doi: 10.1017/S0140525X04000111

Fenk-Oczlon, G., and Fenk, A. (2009a). "Musical pitch in nonsense syllables: correlations with the vowel system and evolutionary perspectives," in *Proceedings of 7th Triennial Conference of the Europaean Society for the Cognitive Sciences of Music*, eds J. Louhivuori, T. Eeerola, S. Saarikallio, T. Himberg, and P.-S. Eerola (Jyväskylä: European Society for the Cognitive Sciences of Music), 110–113.

Fenk-Oczlon, G., and Fenk, A. (2009b). Some parallels between language and music from a cognitive and evolutionary perspective. *Music. Sci.* 13, 201–226. doi: 10.1177/1029864909013002101

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511817779

Fitch, W. T., de Boer, B., Mathur, N., and Ghazanfar, A. A. (2017). Response to Lieberman on "Monkey vocal tracts are speech-ready". *Sci. Adv.* 3:e1701859. doi: 10.1126/sciadv.1701859

Fürniss, S. (1991). *Die Jodeltechnik der Aka-Pygmäen in Zentralafrika*. Berlin: Dieter Reimer.

Gill, K. Z., and Purves, D. (2009). A biological rationale for musical scales. *PLOS ONE* 4:e8144. doi: 10.1371/journal.pone.0008144

Halliday, M. A. K. (1975). *Learning How to Mean: Explorations in the Development of Language*. London: Edward Arnold. doi: 10.1016/B978-0-12-443701-2.50025-1

Hoff, E. (2009). *Language Development*, 4th Edn. Belmont, CA: Wadsworth.

Holzmann, R. (1980). Cuatro ejemplos de música Q'eros (Cuzco, Perú). *Latin Am. Music Rev.* 1, 74–91. doi: 10.2307/780038

Hughes, D. W. (2000). No nonsense: the logic and power of acoustic-iconic mnemonic systems. *Br. J. Ethnomusicol.* 9, 95–122. doi: 10.1080/09681220008567302

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Percept. Interdiscip. J.* 26, 195–204. doi: 10.1525/mp.2009.26.3.195

Jespersen, O. (1894). *Progress in Language*. London: Swan Sonnenschein & Co.

Jordania, J. (2006). *Who Asked the First Question? The Origins of Human Choral Singing, Intelligence, Language and Speech. The Origins of Human Choral Singing, Intelligence*. Tbilisi: Logos.

Kolinsky, R., Pascale Lidji, P., Peretz, I., Besson, M., and Morais, J. (2009). Processing interactions between phonology and melody: vowels sing but consonants speak. *Cognition* 112, 1–20. doi: 10.1016/j.cognition.2009.02.014

Ladefoged, P. (2005). "Speculations on the control of speech," in *Figure of Speech: A Festschrift for John Laver*, eds W. J. Hardcastle and J. MacKenzie Beck (Macwah, NJ: Lawrence Erlbaum), 3–21.

Lauridsen, J. (1983). *Musical Scales in Australian Aboriginal Songs: Structure and Social Implications*. Ph.D. thesis, University of Maryland College Park, College Park, MD.

Lehmann, A. C., and Goldhahn, S. (2016). Duration of playing bursts and redundancy of melodic jazz improvisation in John Coltrane's "Giant Steps". *Music. Sci.* 20, 345–360. doi: 10.1177/1029864916649637

Lewis, J. (2013). "A cross-cultural perspective on the significance of music and dance to culture and society insight from BaYaka pygmies," in *Language, Music and the Brain: A Mysterious Relationship*, ed. M. Arbib (Cambridge: MIT Press), 45–65.

Lidji, P., Jolicoeur, P., Régine Kolinsky, R., Moreau, P., Connolly, J. F., and Peretz, I. (2010). Early integration of vowel and pitch processing: a mismatch negativity study. *Clin. Neurophysiol.* 121, 533–541. doi: 10.1016/j.clinph.2009.12.018

Lieberman, P. (2017). Comment on "Monkey vocal tracts are speech-ready." *Sci. Adv.* 7:e1700442. doi: 10.1126/sciadv.1700442

Liu, H.-M., Kuhl, P. K., and Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Dev. Sci.* 6, F1–F10. doi: 10.1111/1467-7687.00275

Maddieson, I. (2005). "Vowel quality inventories," in *The World Atlas of Language Structures*, eds M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie (Oxford: Oxford University Press), 14–17.

Marks, L. E. (1978). *The Unity of the Senses: Interrelations Among the Modalities*. New York, NY: Academic Press. doi: 10.1016/B978-0-12-472960-5.50011-1

Masataka, N. (2007). Music, evolution and language. *Dev. Sci.* 10, 35–39. doi: 10.1111/j.1467-7687.2007.00561.x

Meyer, E. A. (1896). Zur tonbewegung des vokals im gesprochenen und gesungenen einzelwort. *Phonet. Stud.* 10, 1–21.

Meyer, J. (2008). Typology and acoustic strategies of whistled languages: phonetic comparison and perceptual cues of whistled vowels. *J. Int. Phonetic Assoc.* 38, 69–94. doi: 10.1017/S0025100308003277

Mithen, S. (2005). *The Singing Neanderthals. The Origins of Music, Language, Mind, and Body*. London: Weidenfeld and Nicolson.

Möbius, B. (2003). "Gestalt Psychology meets phonetics – An early experimental study of intrinsic F0 and intensity," in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 2677–2680.

Nattiez, J.-J. (1990). *Music and Discourse: Toward a Semiology of Music*, trans. C. Abbate. Princeton, NJ: Princeton University Press.

Nettl, B. (1956). *Music in Primitive Culture*. Cambridge: Harvard University Press. doi: 10.4159/harvard.9780674863408

Neuhoff, J. G., and Lidji, P. (2014). Language familiarity, expectation, and novice musical rhythm production. *Lang. Speech* 57, 563–572. doi: 10.1177/0023830914520837

Nikolsky, A. (2015). Evolution of tonal organisation in music mirrors symbolic representation of perceptual reality. Part I: prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.01405

Olsen, P. R. (1972). Acculturation in the Eskimo songs of the greenlanders. *Yearb. Int. Folk Music Council* 4, 32–37. doi: 10.2307/767670

Papousek, M. (1992). "Early ontogeny of vocal communication in parent-infant interactions," in *Nonverbal Communication: Comparative and Developmental Aspects*, eds H. Papousek, U. Jurgens, and M. Papousek (New York, NY: CUP), 230–261.

Patel, A. D., and Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition* 87, B35–B45. doi: 10.1016/S0010-0277(02)00187-7

Peterson, G. E., and Barney, H. L. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175–184. doi: 10.1121/1.1906875

Pierce, J. R. (1999). "The nature of musical sound," in *The Psychology of Music*, 2nd Edn, ed. D. Deutsch (San Diego, CA: Academic Press), 1–23.

Pinker, S. (1997). *How the Mind Works*. New York, NY: W. W. Norton.

Pommer, J. (1893). *252 Jodler and Juchezer*. Wien: Rebay & Robitschek.

Rakowski, A. (1999). Perceptual dimensions of pitch and their appearance in the phonological system of music. *Music. Sci.* 3, 23–29. doi: 10.1177/102986499900300102

Rhodes, W. (1954). *Music of the American Indian: Plains: Comanche, Cheyenne, Kiowa, Caddo, Wichita, Pawnee. From the archive of folk culture*. Washington, DC: The Library of Congress, Motion picture, broadcasting and recorded sound division.

Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abey, C. (1997). Major trends in vowel system inventories. *J. Phonet.* 25, 233–253. doi: 10.1006/jpho.1997.0044

Sheikin, Y. (2002). *History of Music Culture of Siberia Peoples: Comparativ-Historic Investigation* [История музыкальной культуры народов Сибири: сравнительно-историческое исследование]. Moscow: Russian Academy of Science [ВосточнаялитератураРАН].

Traunmüller, H. (1986). "Some aspects of the sound of speech sounds," in *The Psychophysics of Speech Perception*, ed. M. E. H. Schouten (Dordrecht: Martinus Nijhoff), 293–305.

Trehub, S. E., Schellenberg, G., and Kamenetzky, S. B. (1999). Infants' and adults' perception of scale structure. *J. Exp. Psychol. Percept. Perform.* 25, 965–975. doi: 10.1037/0096-1523.25.4.965

Walker, R. (1997). Visual metaphors as music notations for sung vowel spectra in different cultures. *J. New Music Res.* 26, 315–345. doi: 10.1080/09298219708570733

Wiora, W. (1962). La musique à l'époque de la peinture paléolithique. *J. Int. Folk Music Counc.* 14, 1–6. doi: 10.2307/835551

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Fenk-Oczlon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

ORIGINAL RESEARCH
published: 25 September 2017
doi: 10.3389/fpsyg.2017.01664

# White Matter Correlates of Musical Anhedonia: Implications for Evolution of Music

*Psyche Loui[1]*, Sean Patterson[1], Matthew E. Sachs[2], Yvonne Leung[1,3], Tima Zeng[1] and Emily Przysinda[1]*

*[1] Music, Imaging and Neural Dynamics Lab, Department of Psychology, Program in Neuroscience and Behavior, Wesleyan University, Middletown, CT, United States, [2] Department of Psychology, Brain and Creativity Institute, University of Southern California, Los Angeles, CA, United States, [3] The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, NSW, Australia*

Recent theoretical advances in the evolution of music posit that affective communication is an evolutionary function of music through which the mind and brain are transformed. A rigorous test of this view should entail examining the neuroanatomical mechanisms for affective communication of music, specifically by comparing individual differences in the general population with a special population who lacks specific affective responses to music. Here we compare white matter connectivity in BW, a case with severe musical anhedonia, with a large sample of control subjects who exhibit normal variability in reward sensitivity to music. We show for the first time that structural connectivity within the reward system can predict individual differences in musical reward in a large population, but specific patterns in connectivity between auditory and reward systems are special in an extreme case of specific musical anhedonia. Results support and extend the Mixed Origins of Music theory by identifying multiple neural pathways through which music might operate as an affective signaling system.

**Keywords: music, evolution, auditory, affective, communication, diffusion tensor imaging**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky,*
*Braavo! Enterprises, United States*

***Reviewed by:***
*Christopher I. Petkov,*
*Newcastle University, United Kingdom*
*John G. Neuhoff,*
*College of Wooster, United States*
*Chris Baber,*
*University of Birmingham,*
*United Kingdom*

****Correspondence:***
*Psyche Loui*
*ploui@wesleyan.edu*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *19 May 2017*
***Accepted:*** *11 September 2017*
***Published:*** *25 September 2017*

***Citation:***
*Loui P, Patterson S, Sachs ME, Leung Y, Zeng T and Przysinda E (2017) White Matter Correlates of Musical Anhedonia: Implications for Evolution of Music. Front. Psychol. 8:1664. doi: 10.3389/fpsyg.2017.01664*

## INTRODUCTION

Music is celebrated and valued in every human culture, yet we know relatively little about why music exists, or what functions music might serve for humankind. The evolutionary function of music has been a subject of debate since Darwinian times (Darwin, 1871). On the one hand, some scholars espouse views that music is an evolutionary byproduct that confers no cognitive advantage, i.e., that music is "auditory cheesecake" (Pinker, 1997). On the other hand, most researchers in the field of music perception and cognition posit that music serves many adaptive functions (Huron, 2001; Honing et al., 2015). For each of these functions, musical sounds function as an auditory channel for interpersonal communication, possibly preceding speech and language (Mithen, 2007). Thus, the need for interpersonal communication through an auditory channel is at the core of evolutionary pressures that are thought to shape music.

This need for interpersonal communication likely changes the cognitive system by virtue of lasting effects that music exerts upon our species. The notion that music is a human invention but transforms our experience, i.e., that music is a transformative technology of the mind (TTM) (Patel, 2008, 2010), is attractive for two reasons. Firstly, the TTM view reconciles the debate between more traditional adaptationist and exaptationist views (cf. Justus and Hutsler, 2005; Trainor, 2006) by pointing out the false dichotomy between these two views. Secondly, TTM brings to the forefront

the idea that brains can change as a result of musical experience. Thus, studies that relate musical experience to inter-individual variability within our species may be informative of how music came to be valued in our species. Another important evolutionary role of music is in its value as an emotional signal: music has power to communicate and evoke strong emotions through an auditory channel (Snowdon et al., 2015). Building on to Patel's TTM theory, the Mixed Origins of Music (MOM) theory posits that music transforms the brain through an affective signaling system that is common to many socially living animals (Altenmüller et al., 2013b; Snowdon et al., 2015). Specifically, the neural mechanisms through which chills occur in response to music may be informative of the evolution of music as an affective communication tool (Altenmüller et al., 2013b). Music elicits a variety of emotions: from abstract, aesthetic experiences to strong, more physiologically measurable emotional responses (Scherer, 2004). Several underlying mechanisms have been proposed through which emotional responses to music might be elicited. For instance, the BRECVEMA model provides a comprehensive account of emotional mechanisms for music (Juslin and Västfjäll, 2008; Juslin, 2013); they include (among others) brainstem responses and evaluative conditioning mechanisms, which involve brain areas within the dopaminergic reward system.

Music that is rewarding is processed by functional connectivity between auditory areas [superior temporal gyrus (STG)] and reward system areas such as the nucleus accumbens (NAcc) (ventral striatum), caudate (dorsal striatum), and areas in the classic limbic system including the amygdala and anterior insula (AIns) (Salimpoor et al., 2011, 2013). Individual differences in the tendency to derive chills, i.e., measurable psychophysiological responses, from music are associated with structural connections from auditory regions (STG) to the AIns, which is consistently activated during the experience of strong emotions, and the medial prefrontal cortex (mPFC), which is important for computing social value; furthermore, this association is modulated by connectivity through the NAcc, a hub in the dopaminergic reward system (Sachs et al., 2016). On the extreme end of the spectrum of individual differences in musical reward, recent work has found evidence for specific musical anhedonia, a rare but intriguing condition where individuals derive no reward responses from their musical experience (Mas-Herrero et al., 2013, 2014). The underlying brain mechanisms are similar to those reviewed above, in that they involve functional connectivity between auditory regions and reward regions, notably the dopaminergic pathway centering around the NAcc (Martínez-Molina et al., 2016).

Although differences in functional connectivity and general brain structure have both been observed in subjects with musical anhedonia (Martínez-Molina et al., 2016; Belfi et al., 2017), it is not yet known whether, and to what extent, these functional connectivity differences identified in musical anhedonics might also be structurally detectable. Furthermore, although structural brain differences in white matter connectivity between auditory and emotion and reward areas have been related to individual differences in reward responses to music (Sachs et al., 2016), it is unknown whether specific musical anhedonia simply reflects the low end of a continuum of normal individual differences in brain connectivity and reward responses to music, or whether musical anhedonia is a categorically distinct disorder that reflects anatomically dissociated neural substrates from normal variations in reward sensitivity. If the former is the case (i.e., musical anhedonia represents the low end of a single continuum), then one would expect that differences in auditory-to-reward connectivity between musical anhedonics also extend to the rest of the population. Conversely, if the latter is the case (i.e., musical anhedonia is different from normal variability in musical reward), one would expect that musical anhedonics have different patterns of reward and auditory-to-reward connectivity from the variations that are generally observed within the population that reflect reported differences in the experience of reward from music.

Here we test the primary hypothesis that musical anhedonia reflects specific differences in white matter connectivity within the reward system, and between the auditory and reward systems. Secondly, we test the hypothesis that the same patterns of white matter connectivity reflect individual differences in the normal variations of reward experiences in music. Using combined behavioral and diffusion tensor imaging (DTI) methods, we compare the white matter connectivity of a musically anhedonic subject, BW, to a group of normal controls ($n = 46$) who report a range of reward from music. Results will identify the neuroanatomical networks that predispose the human brain toward successful affective communication through music.

# MATERIALS AND METHODS

## Subjects

Subject BW (male, age 53 years, right-handed) presented with a self-reported, socially debilitating lack of reward experience from music despite intact reward responses to visual art. **Table 1** shows demographic information and information about musical training. Screening measures including Montreal Battery for Evaluation of Amusia (Peretz et al., 2003) and the nonverbal measure of the Shipley Institute of Living Scale (Shipley, 1940) were used to rule out any differences due to amusia or general intellectual impairment, respectively.

Control subjects ($n = 46$, 17 females, all right-handed) consisted of Wesleyan students and community members. Subjects reported a variety of musical training, and tested within normal ranges for MBEA and Shipley (**Table 1**). Among the control subjects, 85% (39 subjects) completed the BMRQ. All subjects gave written informed consent as approved by the Institutional Review Boards of Wesleyan University and Hartford Hospital.

## Stimuli

In addition to screening tools reported above, 39 of the 46 subjects completed the Revised Physical Anhedonia Scale (PAS) (Chapman et al., 1976) and the Barcelona Music Reward Questionnaire (BMRQ) (Mas-Herrero et al., 2013). The PAS is a self-report scale used to measure anhedonia, the lowered ability to experience pleasure (Chapman et al., 1976). It consists of 61

statements that describe pleasurable experiences (e.g., "I have usually found lovemaking to be intensely pleasurable."). Subjects are asked to indicate whether each statement is true or false as it applies to them. Among the 61 statements, 10 items pertain to sounds (e.g., "The sounds of a parade have never excited me.") whereas the others are non-sound items that include other sensory and social pleasures (e.g., "I have often found walks to be relaxing and enjoyable." "I have often enjoyed receiving a strong, warm handshake.").

The BMRQ (Mas-Herrero et al., 2013) was used to assess how BW experienced reward associated with music, in comparison with the control group. The BMRQ is a 20-item questionnaire designed to measure musical reward experiences as a combination of five factors: musical seeking, emotion evocation, mood regulation, sensory-motor, and social reward.

## Procedures

After informed consent procedures, subjects completed surveys to report their demographic and musical training data. They also completed the MBEA and Shipley tests as screening measures for amusia and intellectual impairment. They then completed the PAS and BMRQ to assess possible general and musical anhedonia.

In addition to behavioral data, high-resolution T1 and DTI images were acquired in a 3T Siemens Skyra MRI scanner at the Olin Neuropsychiatry Research Center at the Institute of Living. Anatomical images were acquired using a T1-weighted, 3D, magnetization-prepared, rapid-acquisition, gradient echo (MPRAGE) volume acquisition with a voxel resolution of 0.8 mm × 0.8 mm × 0.8 mm. Diffusion images were acquired using a diffusion-weighted, spin-echo, echo-planar imaging sequence (TR = 4.77 s, voxel size = 2.0 mm × 2.0 mm × 2.0 mm, axial acquisition, 64 noncollinear directions with *b*-value of 1000 s/mm$^2$, 64 noncollinear directions with *b*-value of 2000 s/mm$^2$, 1 image with *b*-value of 0 s/mm$^2$).

**TABLE 1 |** Demographic information, baseline tests, and scores on Barcelona Music Reward Questionnaire and Physical Anhedonia Scale for BW and control subjects.

| | Control group; mean (SD) | BW |
|---|---|---|
| *N* | 46 | 1 |
| Number of females | 17 | N/A |
| Age | 20.5 (4.66) | 53 |
| Years of musical training | 7.30 (4.44) | 4 |
| Age of onset of musical training | 8.35 (2.98) | 13 |
| Shipley score | 17.2 (1.91) | 15 |
| MBEA (% correct) | 81.4% (6.83%) | 80.6% |
| **BMRQ scores** | | |
| Musical reward | 49.0 (11.3) | −9 |
| Music seeking | 53.4 (11.4) | 13 |
| Emotional evocation | 46.2 (13.3) | 2 |
| Mood regulation | 49.0 (12.2) | −1 |
| Sensorimotor | 44.3 (10.6) | 12 |
| Social reward | 55.4 (11.9) | 24 |
| **PAS (% anhedonic score)** | | |
| Sound items | 19.6% (13.7%) | 90.9% |
| Non-sound items | 16.7% (11.2%) | 21.7% |

## Data Analysis

All MR images were processed using FMRIB's Software Library (FSL) (Jenkinson et al., 2012). The images were then corrected for eddy current distortions using the eddy correct function. Non-brain structures were removed from each participant's images by the brain extraction tool. A diffusion tensor model was fit at each voxel in the extracted brain using the dtifit function to get a fractional anisotropy (FA) image for each participant. Probabilistic tractography was conducted using a Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques (bedpostX) to determine the probable directions of each fiber for each brain voxel (Behrens et al., 2007).

Probabilistic tractography was conducted to determine structural connectivity in each hemisphere between each pair of the following regions of interest: STG, AIns, mPFC, and NAcc. The same regions were used as in our previous study (Sachs et al., 2016), as they were specifically identified to include white matter regions within the reward system (mPFC, NAcc, and aIns) and the auditory system (STG). The STG and NAcc were extracted from the Harvard-Oxford Cortical atlas (Desikan et al., 2006), and masked with a standardized FA image. The AIns was extracted from the LONI atlas (Shattuck et al., 2008). Then, using previous literature as a reference (Uddin and Menon, 2009), the anterior portion was defined anatomically within the lateral sulcus. As atlases varied in their delineation of the prefrontal cortex, the mPFC was hand drawn on coronal slices in the anterior portion of the corona radiata (Marchina et al., 2011). Each ROI was extracted or hand drawn on the standardized FA template by a first coder, and verified by a second coder. Each ROI was then warped to each individual subject's FA image in native space and binarized.

Tractography was then initiated from each ROI as a seed toward each other ROI as a waypoint mask; and then tractography was initiated again using the original waypoint mask as the seed and the original seed as the waypoint mask; these two directions of probabilistic tractography were then averaged to yield a single tract between each pair of regions. Each resultant tract was averaged and then thresholded at 10% of its robust intensity level to minimize extraneous tracts. Tract volume and mean FA of the normalized tracts were exported for statistical comparisons. Additionally, to enable visualization all subjects' tracts and FA images were aligned and normalized to the FSL 1 mm FA template using both linear registration (FLIRT) (Jenkinson and Smith, 2001) and nonlinear registration (FNIRT) tools, and canonical tract images were created by averaging each binarized tract across subjects in the control group, and thresholding voxels below the median.

One-sample *z*-tests were used to compare tract volume and normalized FA between BW and control group. Furthermore, to test for brain–behavior relationships within the control group, we ran two separate multiple regression models, both using Music Reward (overall score from the BMRQ) as the dependent variable. The first regression used FA values from each tract as

predictor variables; the second regression used volumes from each tract as predictor variables. Collinearity for all variables in both regressions was minimal (Tolerance > 0.1, VIF < 8). For tracts that were significant predictors of the Music Reward score, we also conducted follow-up tests for correlations between tract FA and each of the five subscores from BMRQ, while applying the Bonferroni statistical correction for the five subscores.

# RESULTS

## Behavioral Results

Barcelona Music Reward Questionnaire showed that BW had low reward response to music in all categories of musical reward. While controls had an average factor score of 50 ($SD = 10$) on the BMRQ (Music Reward overall score), BW had an overall factor score of −9, which was 5.89 standard deviations below controls. BW scored more than 2.5 standard deviations below controls on all subscales of the BMRQ (**Figure 1A**).

Physical Anhedonia Scale showed that BW was not generally anhedonic, except for items that pertain to sound. Control subjects generally scored an average of 17% of responses in the anhedonic ("pathological") direction ($SD = 9\%$). BW scored a total of 39% of responses in the anhedonic direction. Item analysis of the PAS was done by separately analyzing sound and non-sound categories. While the control subjects showed similar proportions of anhedonic scores for sound items and non-sound items ($M = 19.6\%$, $SD = 14\%$ anhedonic responses for sound items; $M = 16.8\%$, $SD = 11\%$ anhedonic responses for non-sound items); BW showed 21.7% anhedonic scores for non-sound items (within 1 SD of the mean) but 90.9% anhedonic responses for sound items (more than 5 SD above the mean). This striking dissociation (**Figure 1B**) suggests that BW does not have general anhedonia, but is specifically anhedonic toward sounds, especially to music.

## DTI Results

### Musical Anhedonic vs. Controls

**Figure 2** compares tract FA and volume between BW and control subjects, showing some differences in auditory–reward connectivity in the subject with musical anhedonia. BW had significantly lower tract volume than controls in tracts between the left STG and left NAcc ($z = -2.16$, $p = 0.03$) and between the left AIns and left NAcc ($z = -1.98$, $p = 0.04$) at the uncorrected $p < 0.05$ level. No other tracts showed statistically significant differences between BW and controls according to $z$-tests. Mean FA (after normalizing for volume to enable a direct comparison of FA values) was greater for BW than controls between left STG and left AIns ($z = 3.08$, $p = 0.002$), the same tract in which he showed lower volume than controls, surviving Bonferroni correction at $p < 0.05/10$. No other tracts showed significant differences in FA according to the $z$-test.

### Individual Differences within Control Group

A multiple regression model with the dependent variable of Music Reward score, with tract volume (in $mm^3$) of each tract as predictor variables, accounted for 38% of the variability ($R^2 = 0.38$), but was not significant after accounting for the number of predictors (adjusted $R^2 = 0.15$, $F = 1.69$, $p = 0.13$). Among the controls, the Music Reward score was significantly predicted by the volume of tracts between LSTG and LAIns ($\beta = 1.11$, $t = 2.76$, $p = 0.01$, bivariate correlation $r = 0.26$, partial correlation $r_p = 0.463$), between RSTG and RNAcc ($\beta = -0.81$, $t = -2.33$, $p = 0.027$, $r = 0.036$, $r_p = -0.40$), and between RSTG and RMPFC ($\beta = 0.74$, $t = 2.10$, $p = 0.045$, $r = 0.193$, $r_p = 0.37$). Although these tract volumes were significant predictors of Music Reward at the $p < 0.05$ level, they did not survive correction for multiple comparisons across the 10 tested tracts. **Figure 3** shows these tracts and scatterplots of their bivariate correlations with the Music Reward score.

A multiple regression model with the dependent variable of Music Reward score, with FA values of each tract as predictor variables, accounted for 26% of the variability ($R^2 = 0.26$, adjusted $R^2 = -0.002$, $F = 0.99$, $p = 0.47$). None of the tracts emerged as significant predictors (all $p > 0.05$).

### Predicting Musical Anhedonic Brain and Behavior from Control Group Data

To assess whether BW falls along the same continuum of brain–behavior relationships as predicted by controls, we first used the regression model from all tract volume data to generate a prediction for BW's Music Reward score. Given the multiple regression model obtained from tract volume data above (see section "Individual Differences Within Control Group"), BW's tract volume data predicted his Music Reward score to be 0.29, which was much higher than his actual score (−9). However, pairwise correlations between behavior and tract volume (scatterplots in **Figure 3**) showed that BW is a predictable outlier from the control subjects' data, with low volume in tracts between LSTG and LAIns and between RSTG and RMPFC, as predicted by his low Music Reward score and by control subjects' data. To assess whether BW's tract volumes belonged to the same continuum as controls, we used the slope and intercept of the trend line that best fit the bivariate relationship among control subjects to predict BW's tract volumes using his Music Reward score (**Table 2**), thus extrapolating control subjects' data to predict BW's tract volumes. The prediction fits BW's actual data with 7.6% error for LSTG_LAIns tract, with 9.4% error for the RSTG_RNAcc tract, and with 1.0% error for the RSTG_RMPFC tract, suggesting that for these three tracts, BW falls on the extreme end of the same continuum as the control subjects.

# DISCUSSION

Individual differences in brain and behavior can be demonstrated by the normal variance within the general population, as well as extreme cases where substantial variations in brain and behavior give rise to striking deviations from the general population. To the evolution of music, the existence of musical anhedonia presents one such model of a striking dissociation, in which some individuals have a lack of reward responses specifically to sound. Here, we see that patterns of white matter connectivity in the auditory and reward systems reflect individual differences in

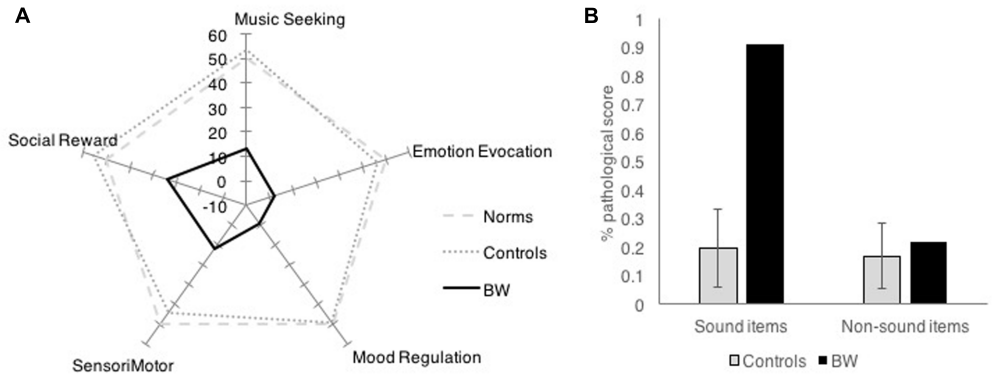



**FIGURE 1 |** Survey results comparing BW against controls in **(A)** Barcelona Music Reward Questionnaire and **(B)** Physical Anhedonia Scale.

the tendency to perceive reward from music. Auditory–reward connectivity differences are observed in our extreme case of musical anhedonia, and also reflect individual differences in music reward sensitivity within the control group.

BW, a subject with severe musical anhedonia, had decreased white matter volume but higher FA between auditory and reward areas, specifically between left STG and left NAcc. The left STG is a cortical hub of the auditory system: it includes auditory belt and parabelt areas which are important for analyzing temporal content of sounds, including speech-specific content (Overath et al., 2008, 2015). The NAcc is central to the mesolimbic pathway of the dopaminergic reward system, with its known role in reward and reinforcement (Wise, 2006), and is the crucial waystation of a reward network activated during the peak experience of music-related reward (Salimpoor et al., 2011; Zatorre and Salimpoor, 2013; Koelsch, 2014). Although the left NAcc showed higher volume of connectivity to the ipsilateral AIns as well as STG, the volume results were only significant at the uncorrected level; in contrast, the increased FA between left NAcc and STG was significant at the Bonferroni-corrected level in BW. FA, the main outcome variable in DTI, is an index of white matter integrity which includes myelination and coherence of axonal bundles. Probabilistic tractography requires FA values of each voxel to be above the white matter threshold, in order to derive tract volume (Behrens et al., 2007). Here, the pattern of simultaneously increased white matter integrity and decreased volume may suggest increased myelination and/or decreased crossing fibers in BW's anatomical connections between LSTG and LNAcc, which could result in increased inhibition from LSTG to LNAcc. Functionally, the increased inhibition from LSTG could lead to a downregulation of the activity of LNAcc, resulting in deactivation of the NAcc as observed in recent functional MRI work in musical anhedonics (Martínez-Molina et al., 2016). Although these results are correlative rather than causal, the finding that BW had decreased volume but increased white matter integrity between these two regions adds to existing literature on the role of auditory–reward connectivity in affective responses for music (Salimpoor et al., 2013; Sachs et al., 2016); the implications of this data pattern for the evolution of music will be considered again later in this section "Discussion".

The PAS showed that BW was anhedonic to all sound items, including non-music items (e.g., "the sounds of a parade"; "the cackling of fire in a fireplace"). Upon further interview, BW stated: "The crackle of a fireplace, the rustle of leaves, the swish of ocean waves – I just don't appreciate them." It remains to be seen whether musical anhedonics in other studies also report anhedonia toward sound items from the PAS, or whether BW is unique in his lack of appreciation of all auditory stimuli. If BW is different from other musical anhedonics in this regard, then one might expect that his auditory–reward disconnection is also more general than other cases of musical anhedonia.

Regarding the lack of appreciation for sounds, an interesting related question concerns whether BW could have misophonia, another auditory disorder where an individual reacts aversively to trigger sounds (Kumar et al., 2017). While more research is needed in the future to determine the extent of overlap or shared traits between misophonia and musical anhedonia, our study identifies BW as having musical anhedonia rather than misophonia, mainly because BW's main complaint is that he feels no enjoyment from music, rather than being angered or anxious in response to trigger sounds as is common among misophonics (Edelstein et al., 2013). According to his self-report: "Music doesn't particularly change my mood or give me an emotional response." "Music never disgusts me. (The taste of cheese disgusts me. The smell of rotten eggs disgusts me. The sight of gore disgusts me.) Mostly I'd say that I'm neutral about music, because I just don't care (and I don't care that I don't care!), and I mostly tune it out." He also reports normal responses to speech and nonverbal vocal sounds. In contrast, misophonics most commonly report feeling disgusted, trapped, and/or anxious in response to trigger sounds, which are typically sounds produced by other people (Edelstein et al., 2013). From our findings, BW shows an abnormal pattern of connectivity from the NAcc; this was not observed in misophonics (Kumar et al., 2017). Thus, at present results suggest that musical anhedonia pertains more to a lack of reward, whereas misophonia pertains more to the experience of negative emotions such as anger and irritation in reaction to trigger sounds.

Within our control group, volume of some tracts between auditory and reward regions, specifically between LSTG and

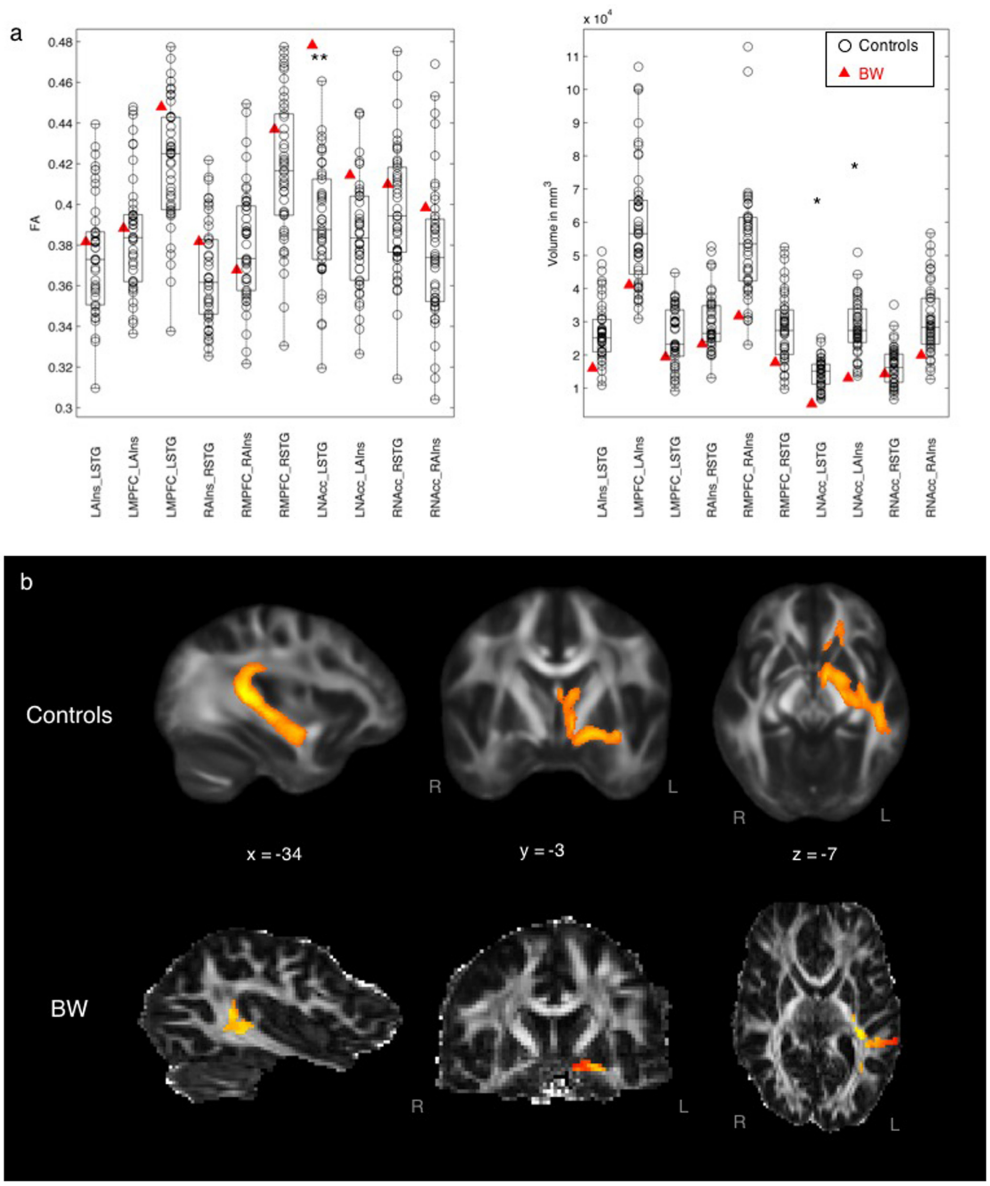



**FIGURE 2 | (a)** Left: Mean FA of each tract comparing BW and controls, controlling for volume differences. $^{**}p < 0.05$ Bonferroni-corrected (0.05/10 = 0.005). Right: Volume of each tract comparing BW and controls. $^{*}p < 0.05$ uncorrected. **(b)** Averaged tract between STG and NAcc for controls (top row) and for BW (bottom row).

LAIns, between RSTG and RNAcc, and between RSTG and RMPFC, were predictive of musical reward at the 0.05 (uncorrected) level. Although these results do not survive correction for multiple comparisons, it is noteworthy that only tracts from left or right STG (the only auditory regions in our model) emerged as significant predictors, suggesting that individual differences in music reward do pertain to auditory-specific access to the reward system. It is also noteworthy that BW's tract volume data can be predicted by extrapolating the trend line that best fits the bivariate relationship between music reward and volume of the significant predictor tracts. In contrast, the multiple regression model obtained from control subjects did not accurately predict BW's music reward score. Thus, control subjects' data can predict BW's tract volumes but not his behavioral scores. This may be because BW's music reward score, at 5.89 *SD* below controls, is much more of an outlier than his brain measures; thus, the brain predictors of behavior derived from control subjects do not apply to BW's very unusual behavioral data, but BW's tract volume data appear to lie at the low end of a normal distribution. The fact that BW is a very extreme outlier on the BMRQ also suggests that true musical anhedonia, at least as represented by the case of BW, is probably very rare. This is consistent with the observation that across patients of many types of brain damage, few report

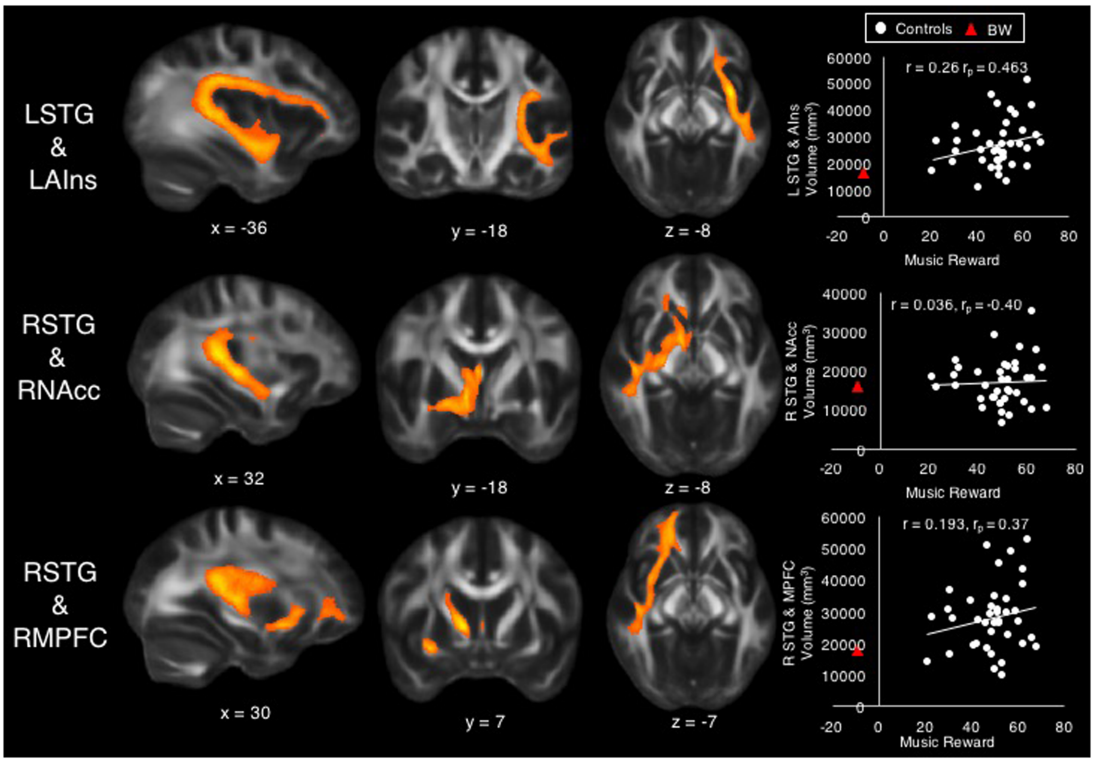



**FIGURE 3 |** Volumes of tracts between left superior temporal gyrus (STG) and left anterior insula (top), between right STG and right nucleus accumbens (middle), and between right STG and right medial prefrontal cortex (bottom) were significant predictors of the Music Reward score among control subjects. Scatterplots show bivariate correlations (*r*) between Music Reward score and the volume of each tract, as well as partial correlation coefficients ($r_p$) from the regression for purposes of comparison against bivariate correlations. BW's data are also shown on scatterplots for purposes of comparison.

musical anhedonia (Belfi et al., 2017). Future studies might rely on more targeted strategies to identify more such cases of musical anhedonia.

The tract between LSTG_LAIns shows a continuum in volume that best reflects our range of behavioral data: its volume is reduced in the musical anhedonic as well as positively correlated with music reward. Connections between AIns and STG likely include the arcuate fasciculus, part of the auditory dorsal pathway that connects superior temporal and inferior frontal regions that is related to musical ability (Loui et al., 2009, 2011; Halwani et al., 2011; Loui, 2015). Furthermore, AIns is reduced in functional connectivity to auditory cortex in singers (Kleber et al., 2013), and functional connectivity between LSTG and LAIns is correlated with lexical retrieval in spontaneous speech (Chai et al., 2016). In addition to its role in vocal–motor integration and speech, the AIns is part of the classic limbic system and is implicated in the quartet theory of emotions due to its importance in interoception and emotional regulation (Koelsch et al., 2015). Given these diverse roles of AIns in the auditory–motor system, the present finding of increased tract volume between left AIns and LSTG in controls who experience high musical reward may relate to auditory–motor behavior especially as it applies to vocal–motor behavior. This auditory–insula connectivity may be related to the differentiation of vocalization repertoire as posited in the MOM theory (Altenmüller et al., 2013a). The MOM theory states that differentiation of vocalization repertoire, as driven by chill experiences, led to the capacity for fine-grained rhythmic–melodic discrimination. In our evolutionary history, it is possible that individuals with high LSTG_LAIns connectivity, who were highly reward-sensitive to music (e.g., frequently experiencing chills in response to music), then went on to acquire fine-grained auditory discrimination skills, which then gave rise to language and music. Since the AIns is an evolutionarily older part of the brain than its neighbor the inferior frontal gyrus (which is a classic endpoint of the arcuate fasciculus) (Galaburda and Pandya, 1982; Semendeferi and Damasio, 2000), the LSTG_LAIns connection could have predated the arcuate fasciculus, thus serving as a pathway for the differentiation of vocalization response that preceded the hypothesized bifurcation of auditory information into music and language (Mithen, 2007).

Superior temporal gyrus connections to NAcc and mPFC may include the arcuate as well as the uncinate fasciculus, the latter being part of the auditory ventral pathway that connects the temporal and frontal lobes (Wakana et al., 2004) and is involved in processing local syntactic structures (Friederici, 2009). mPFC is also part of the default mode network and is involved in social, self-referential, and emotional processing (Fox et al., 2005; Mason et al., 2007; Jenkins and Mitchell, 2010; Kim and Johnson, 2015). As the mPFC is a waystation of the dopaminergic system that probably emerged later in evolution (Galaburda

**TABLE 2 |** Predicting musical anhedonic from control data.

| Tract | Slope | Intercept | Prediction | Actual | Error[a] |
|---|---|---|---|---|---|
| Predicting BW's music reward score from overall regression model | | | | | |
| All | 5.23 | 47.36 | 43.58 | −9 | 5.84 |
| Predicting BW's tract volume scores from bivariate correlation trend lines | | | | | |
| RSTG_RMPFC | 178.49 | 19,204 | 17,597.59 | 17,776 | 0.010036566 |
| LSTG_LAIns | 206.02 | 16,768 | 14,913.82 | 16,040 | 0.070210723 |
| RSTG_RNAcc | 18.791 | 15,935 | 15,765.881 | 14,280 | −0.1040533 |

*BW's music reward score was not successfully predicted (103% error) from the overall regression model of DTI predictors. In contrast, BW's tract volume of RSTG_RMPFC, LSTG_LAIns, and RSTG_RNAcc tracts was successfully predicted (10% error) from bivariate correlations between music reward and tract volume.* [a]*[(Actual−Prediction)/Actual].*

and Pandya, 1982; Semendeferi and Damasio, 2000), the finding that connections to it correlate with musical reward suggests a further involvement of an evolutionarily younger part of the dopaminergic system in music processing beyond the NAcc. Interestingly, while the LSTG_LAIns and LSTG_LMPFC tracts show positive bivariate as well as significantly positive partial correlations to music reward, the RSTG_RNAcc tracts show no significant bivariate correlation with music reward, but a significant negative partial correlation after partialling out the effects of the other predictors. This is especially intriguing when considered alongside data from the musical anhedonic subject: BW had a lower volume but higher FA in LSTG_LNAcc; highly hedonic controls had lower volume in RSTG_RNAcc. Together these results suggest that auditory access to the mesolimbic pathway is hemispherically asymmetric, with normal variations in reward sensitivity occurring on the right but abnormal lack of reward on the left. This is consistent with hemispheric asymmetry to attractive vs. aversive stimuli in animals, but only in learned responses (Besson and Louilot, 1995; Molochnikov and Cohen, 2014). In light of the MOM theory, which posits that chill responses were initially a reward to novel auditory patterns prior to its driving of differentiated vocalization repertoire as discussed above, the present findings link the STG_NAcc pathway to this very early step in the evolution of music.

While this study cannot tease apart when or how these individual differences emerged, the pattern of results can be considered in the context of known steps in brain evolution as well as development, which together provide support for the MOM theory. Our rare case of musical anhedonia possesses a different pattern of white matter pathways between auditory regions and reward-sensitive regions, possibly due to abnormal neuronal migration *in utero* or early in development. In the multiple regression analysis to predict musical reward scores from diffusion measures, since we tested pairwise connections between regions in the auditory and reward networks, this necessarily resulted in an elevated number of statistical comparisons. The brain–behavior relationships within the control group are only significant at the uncorrected level. Thus, although the current results are interesting they should be interpreted cautiously until further verification. Nevertheless, the FA difference between BW and the control in the LSTG–NAcc tract survives correction for comparisons across the 10 tested tracts; this gives us higher confidence in a structural difference between auditory and reward areas that is linked to musical anhedonia.

A remaining question concerns whether musical anhedonia is likely to be a spectrum disorder. The answer to this question depends on how we define musical anhedonia. Considering that the BMRQ is for now the only diagnostic tool explicitly in use to identify musical anhedonia, and it yields a continuum of scores when administered to a large population (Mas-Herrero et al., 2013), the lack of musical reward appears to be continuously distributed. On the other hand, if we define musical anhedonia by self-identification of a socially debilitating lack of reward experiences specific to music, then it might not be a spectrum. However, defining musical anhedonia by self-identification would mean that identification depends upon the subject's awareness of their own condition, which would in turn depend on their social environment. For instance, if BW had not heard about musical anhedonia, or if he lived in an environment where music was less celebrated, then he might not have become aware of his condition. Thus, large-scale testing of musical reward sensitivity across different cultures may be helpful for future definitions of cultural norms against which we define musical anhedonia.

Results show that musical anhedonia is related to different patterns of connectivity from auditory to emotion and reward centers of the brain. This auditory access to the reward system informs the evolutionary basis of music: perhaps music evolved as a direct auditory pathway toward social and emotional reward centers in the brain.

With regard to the shared evolutionary basis of music with language, it is worth noting that in contrast to music, language does not seem to achieve the same set of evolutionary functions; that is, although language and music both involve connectivity between auditory, motor, and cognitive systems, language has more direct and specific sound-to-meaning mappings, but music more readily establishes aesthetic or emotional connections such as chills (Silvia and Nusbaum, 2011). Thus, language and music may have shared evolutionary origins as a protolanguage (Mithen, 2007), but their divergence led to different evolutionary functions and outcomes.

Successful musical communication depends on an auditory channel through which reward and emotional areas can be accessed. This is consistent with views of music as mixed origins, which posits that music evolved from evolutionarily ancient chill

reactions to affiliative sounds (Altenmüller et al., 2013b) that then transform the mind (Patel, 2008). Evolutionarily, the emotional content of sound might have accessed these auditory–reward pathways, which then predisposed the brain toward developing reward sensitivity and thus the need for successful emotional communication. In that regard, results suggest that other species who have connectivity between auditory and reward systems would also be able to enjoy music given the appropriate exposure.

Previous work on congenital amusia has been discussed in terms of its implications on the evolution of music (Patel, 2008); in particular white matter connectivity in congenital amusia supports the hypothesis for a shared basis of music and language (Loui et al., 2009; Loui, 2015). Similarly, white matter connectivity in musical anhedonia informs the evolutionary basis of music on emotion. While reward pathways and auditory perception–action pathways are conventionally seen as separate and dissociable systems in the brain, the present study suggests that they operate in concert, and that this concert of brain systems may be important for the evolution of music: in fact, they may provide support for the MOM as tools to transform the mind (Kleinman, 2015).

Individual differences in structural connectivity between the auditory and reward networks likely represent normal variation in musical reward sensitivity, with some additional patterns that give rise to extreme cases such as musical anhedonia. While increased connectivity between auditory and reward networks is indicative of intense emotional responses to music such as frissons (Harrison and Loui, 2014; Sachs et al., 2016), decreased volume coupled with increased myelination or coherence between specific nodes of these networks reflects the striking lack of specific emotional responses as observed in musical anhedonia. By distinguishing between common variations and rare extremes in individual differences in musical reward sensitivity, the present study attempts to extend the MOM theory by identifying distinct neural pathways through which music might operate as an affective signaling system.

## AUTHOR CONTRIBUTIONS

PL conceptualized the idea, designed the study, and wrote the first draft of the manuscript. SP, MS, and EP acquired and analyzed the data. YL and TZ contributed DTI data analyses. All authors revised the manuscript and approved the submission.

## FUNDING

This work was supported by the Imagination Institute, Grammy Foundation, and NSF STTR grant 1720698 to PL.

## ACKNOWLEDGMENT

We thank BW and our control subjects for participating in the study.

## REFERENCES

Altenmüller, E., Kopiez, R., and Grewe, O. (2013a). "A contribution to the evolutionary basis of music: lessons from the chill response," in *The Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford: Oxford University Press), 313–335.

Altenmüller, E., Kopiez, R., and Grewe, O. (2013b). "Strong emotions in music: Are they an evolutionary adaptation?," in *Sound - Perception - Performance*, Vol. 1, ed. R. Bader (Cham: Springer International Publishing), 131–156.

Behrens, T. E., Johansen-Berg, H., Jbabdi, S., Rushworth, M. F., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *Neuroimage* 34, 144–155.

Belfi, A. M., Evans, E., Heskje, J., Bruss, J., and Tranel, D. (2017). Musical anhedonia after focal brain damage. *Neuropsychologia* 97, 29–37. doi: 10.1016/j.neuropsychologia.2017.01.030

Besson, C., and Louilot, A. (1995). Asymmetrical involvement of mesolimbic dopaminergic neurons in affective perception. *Neuroscience* 68, 963–968. doi: 10.1016/0306-4522(95)00255-H

Chai, X. J., Berken, J. A., Barbeau, E. B., Soles, J., Callahan, M., Chen, J. K., et al. (2016). Intrinsic functional connectivity in the adult brain and success in second-language learning. *J. Neurosci.* 36, 755–761. doi: 10.1523/JNEUROSCI.2234-15.2016

Chapman, L. J., Chapman, J. P., and Raulin, M. L. (1976). Scales for physical and social anhedonia. *J. Abnorm. Psychol.* 85, 374–382. doi: 10.1037/0021-843X.85.4.374

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Edelstein, M., Brang, D., Rouw, R., and Ramachandran, V. S. (2013). Misophonia: physiological investigations and case descriptions. *Front. Hum. Neurosci.* 7:296. doi: 10.3389/fnhum.2013.00296

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci.* 13, 175–181. doi: 10.1016/j.tics.2009.01.001

Galaburda, A. M., and Pandya, D. N. (1982). "Role of architectonics and connections in the study of primate brain evolution," in *Primate Brain Evolution*, eds E. Armstrong and D. Falk (Boston, MA: Springer), 203–216.

Halwani, G. F., Loui, P., Rueber, T., and Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. *Front. Psychol.* 2:156. doi: 10.3389/fpsyg.2011.00156

Harrison, L. D., and Loui, P. (2014). Thrills, chills, frissons, and skin orgasms: toward an integrative model of transcendent psychophysiological experiences in music. *Front. Psychol.* 5:790. doi: 10.3389/fpsyg.2014.00790

Honing, H., ten Cate, C., Peretz, I., and Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philos. Trans R. Soc. B Biol. Sci.* 370:20140088. doi: 10.1098/rstb.2014.0088

Huron, D. (2001). Is music an evolutionary adaptation? *Ann. N. Y. Acad. Sci.* 930, 43–61. doi: 10.1111/j.1749-6632.2001.tb05724.x

Jenkins, A. C., and Mitchell, J. P. (2010). Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cereb. Cortex* 20, 404–410. doi: 10.1093/cercor/bhp109

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–156. doi: 10.1016/S1361-8415(01)00036-6

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* 10, 235–266. doi: 10.1016/j.plrev.2013.05.008

Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–575. doi: 10.1017/S0140525X08005293

Justus, T., and Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: assessing innateness and domain specificity. *Music Percept.* 23, 1–27. doi: 10.1525/mp.2005.23.1.1

Kim, K., and Johnson, M. K. (2015). Activity in ventromedial prefrontal cortex during self-related processing: Positive subjective value or personal significance? *Soc. Cogn. Affect. Neurosci.* 10, 494–500. doi: 10.1093/scan/nsu078

Kleber, B., Zeitouni, A. G., Friberg, A., and Zatorre, R. J. (2013). Experience-dependent modulation of feedback integration during singing: role of the right anterior insula. *J. Neurosci.* 33, 6070–6080. doi: 10.1523/JNEUROSCI.4418-12.2013

Kleinman, K. (2015). Darwin and Spencer on the origin of music: Is music the food of love? *Prog. Brain Res.* 217, 3–15. doi: 10.1016/bs.pbr.2014.11.018

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Koelsch, S., Jacobs, A. M., Menninghaus, W., Liebal, K., Klann-Delius, G., von Scheve, C., et al. (2015). The quartet theory of human emotions: an integrative and neurofunctional model. *Phys. Life Rev.* 13, 1–27. doi: 10.1016/j.plrev.2015.03.001

Kumar, S., Tansley-Hancock, O., Sedley, W., Winston, J. S., Callaghan, M. F., Allen, M., et al. (2017). The brain basis for misophonia. *Curr. Biol.* 27, 527–533. doi: 10.1016/j.cub.2016.12.048

Loui, P. (2015). A dual-stream neuroanatomy of singing. *Music Percept.* 32, 232–241. doi: 10.1525/mp.2015.32.3.232

Loui, P., Alsop, D., and Schlaug, G. (2009). Tone deafness: a new disconnection syndrome? *J. Neurosci.* 29, 10215–10220. doi: 10.1523/JNEUROSCI.1701-09.2009

Loui, P., Li, H. C., and Schlaug, G. (2011). White matter integrity in right hemisphere predicts pitch-related grammar learning. *Neuroimage* 55, 500–507. doi: 10.1016/j.neuroimage.2010.12.022

Marchina, S., Zhu, L. L., Norton, A., Zipse, L., Wan, C. Y., and Schlaug, G. (2011). Impairment of speech production predicted by lesion load of the left arcuate fasciculus. *Stroke* 42, 2251–2256. doi: 10.1161/STROKEAHA.110.606103

Martínez-Molina, N., Mas-Herrero, E., Rodriguez-Fornells, A., Zatorre, R. J., and Marco-Pallares, J. (2016). Neural correlates of specific musical anhedonia. *Proc. Natl. Acad. Sci. U.S.A.* 113, E7337–E7345. doi: 10.1073/pnas.1611211113

Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., and Rodriguez-Fornells, A. (2013). Individual differences in Music Reward experiences. *Music Percept.* 31, 118–138. doi: 10.1525/mp.2013.31.2.118

Mas-Herrero, E., Zatorre, R. J., Rodriguez-Fornells, A., and Marco-Pallares, J. (2014). Dissociation between musical and monetary reward responses in specific musical anhedonia. *Curr. Biol.* 24, 699–704. doi: 10.1016/j.cub.2014.01.068

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., and Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science* 315, 393–395. doi: 10.1126/science.1131295

Mithen, S. J. (2007). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. Cambridge, MA: Harvard University Press.

Molochnikov, I., and Cohen, D. (2014). Hemispheric differences in the mesostriatal dopaminergic system. *Front. Syst. Neurosci.* 8:110. doi: 10.3389/fnsys.2014.00110

Overath, T., Kumar, S., von Kriegstein, K., and Griffiths, T. D. (2008). Encoding of spectral correlation over time in auditory cortex. *J. Neurosci.* 28, 13268–13273. doi: 10.1523/JNEUROSCI.4596-08.2008

Overath, T., McDermott, J. H., Zarate, J. M., and Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nat. Neurosci.* 18, 903–911. doi: 10.1038/nn.4021

Patel, A. D. (2008). Music as a transformative technology of the mind. *Paper Presented at Music: Its Evolution, Cognitive Basis, and Spiritual Dimensions,* Cambridge, 18–20.

Patel, A. D. (2010). "Music, biological evolution, and the brain," in *Emerging Disciplines: Shaping New Fields of Scholarly Inquiry in and Beyond the Humanities*, eds C. Levander and C. Henry (Houston, TX: Rice University Press), 91–144.

Peretz, I., Champod, A. S., and Hyde, K. (2003). Varieties of musical disorders. The montreal battery of evaluation of amusia. *Ann. N. Y. Acad. Sci.* 999, 58–75. doi: 10.1196/annals.1284.006

Pinker, S. (1997). *How the Mind Works*. New York, NY: Norton.

Sachs, M. E., Ellis, R. J., Schlaug, G., and Loui, P. (2016). Brain connectivity reflects human aesthetic responses to music. *Soc. Cogn. Affect. Neurosci.* 11, 884–891. doi: 10.1093/scan/nsw009

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat. Neurosci.* 14, 257–262. doi: 10.1038/nn.2726

Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., and Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219. doi: 10.1126/science.1231059

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *J. New Music Res.* 33, 239–251. doi: 10.1080/0929821042000317822

Semendeferi, K., and Damasio, H. (2000). The brain and its main anatomical subdivisions in living hominoids using magnetic resonance imaging. *J. Hum. Evol.* 38, 317–332. doi: 10.1006/jhev.1999.0381

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., et al. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 39, 1064–1080. doi: 10.1016/j.neuroimage.2007.09.031

Shipley, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.* 9, 371–377. doi: 10.1080/00223980.1940.9917704

Silvia, P. J., and Nusbaum, E. C. (2011). On personality and piloerection: individual differences in aesthetic chills and other unusual aesthetic experiences. *Psychol. Aesthet. Creat. Arts* 5, 208–214. doi: 10.1371/journal.pone.0130117

Snowdon, C. T., Zimmermann, E., and Altenmuller, E. (2015). Music evolution and neuroscience. *Prog. Brain Res.* 217, 17–34. doi: 10.1016/bs.pbr.2014.11.019

Trainor, L. J. (2006). Innateness, learning, and the difficulty of determining whether music is an evolutionary adaptation. *Music Percept.* 24, 105–110. doi: 10.1525/mp.2006.24.1.105

Uddin, L. Q., and Menon, V. (2009). The anterior insula in autism: under-connected and under-examined. *Neurosci. Biobehav. Rev.* 33, 1198–1203. doi: 10.1016/j.neubiorev.2009.06.002

Wakana, S., Jiang, H., Nagae-Poetscher, L. M., van Zijl, P. C., and Mori, S. (2004). Fiber tract-based atlas of human white matter anatomy. *Radiology* 230, 77–87. doi: 10.1148/radiol.2301021640

Wise, R. A. (2006). Role of brain dopamine in food reward and reinforcement. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1149–1158. doi: 10.1098/rstb.2006.1854

Zatorre, R. J., and Salimpoor, V. N. (2013). From perception to pleasure: music and its neural substrates. *Proc. Natl. Acad. Sci. U.S.A.* 110(Suppl. 2), 10430–10437. doi: 10.1073/pnas.1301228110

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Loui, Patterson, Sachs, Leung, Zeng and Przysinda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Neuroscience

REVIEW
published: 29 September 2017
doi: 10.3389/fnins.2017.00519

# Evolutionary Musicology Meets Embodied Cognition: Biocultural Coevolution and the Enactive Origins of Human Musicality

*Dylan van der Schyff*[1, 2*] *and Andrea Schiavio*[3, 4, 5]

[1] Faculty of Education, Simon Fraser University, Burnaby, BC, Canada, [2] Faculty of Music, University of Oxford, Oxford, United Kingdom, [3] Institute for Music Education, University of Music and Performing Arts, Graz, Austria, [4] Department of Music, The University of Sheffield, Sheffield, United Kingdom, [5] Centre for Systematic Musicology, University of Graz, Graz, Austria

Despite evolutionary musicology's interdisciplinary nature, and the diverse methods it employs, the field has nevertheless tended to divide into two main positions. Some argue that music should be understood as a naturally selected adaptation, while others claim that music is a product of culture with little or no relevance for the survival of the species. We review these arguments, suggesting that while interesting and well-reasoned positions have been offered on both sides of the debate, the nature-or-culture (or adaptation vs. non-adaptation) assumptions that have traditionally driven the discussion have resulted in a problematic *either/or* dichotomy. We then consider an alternative "biocultural" proposal that appears to offer a way forward. As we discuss, this approach draws on a range of research in theoretical biology, archeology, neuroscience, embodied and ecological cognition, and dynamical systems theory (DST), positing a more integrated model that sees biological and cultural dimensions as aspects of the same evolving system. Following this, we outline the enactive approach to cognition, discussing the ways it aligns with the biocultural perspective. Put simply, the enactive approach posits a deep continuity between mind and life, where cognitive processes are explored in terms of how self-organizing living systems enact relationships with the environment that are relevant to their survival and well-being. It highlights the embodied and ecologically situated nature of living agents, as well as the active role they play in their own developmental processes. Importantly, the enactive approach sees cognitive and evolutionary processes as driven by a range of interacting factors, including the socio-cultural forms of activity that characterize the lives of more complex creatures such as ourselves. We offer some suggestions for how this approach might enhance and extend the biocultural model. To conclude we briefly consider the implications of this approach for practical areas such as music education.

**Keywords: origins of music, biocultural coevolution, music cognition, enactive cognition, dynamical systems theory**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky, Braavo! Enterprises, United States*

***Reviewed by:***
*L. Robert Slevc, University of Maryland, College Park, United States*
*Tom Froese, National Autonomous University of Mexico, Mexico*

***Correspondence:**
*Dylan van der Schyff*
*dbvanderschyff@gmail.com*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *23 February 2017*
***Accepted:*** *04 September 2017*
***Published:*** *29 September 2017*

***Citation:***
*van der Schyff D and Schiavio A (2017) Evolutionary Musicology Meets Embodied Cognition: Biocultural Coevolution and the Enactive Origins of Human Musicality. Front. Neurosci. 11:519. doi: 10.3389/fnins.2017.00519*

## INTRODUCTION

The debate over the origins and meaning of music for the human animal is one of the most fascinating areas of inquiry across the sciences and humanities. Despite the diversity of perspectives on offer, however, this field has traditionally been guided by approaches that see adaptation by natural selection as the central mechanism driving evolutionary processes (Huron, 2001; for a discussion see Tomlinson, 2015). This extends to the brain, which is often understood as a computing machine that evolved to solve the kinds of problems faced by our prehistoric ancestors in their everyday lives (see Anderson, 2014). Importantly, this "adaptationist" orientation posits a rather strict separation between the products of natural selection (i.e., adaptations) and those of culture. Because of this, evolutionary musicologists have often been faced with something of a dichotomy: Music tends to be seen *either* as a naturally selected adaptation that has contributed directly to our survival as a species, *or* as a product of culture with little or no direct connection to our biological heritage (see van der Schyff, 2013a; Tomlinson, 2015; Killin, 2016a, 2017). Various arguments have emerged in support of each position (more on this below; see Pinker, 1997; Huron, 2001; Mithen, 2005; Patel, 2008; Honing et al., 2015). Moreover, the influence of the computational model of mind has tended to focus research and theory in music cognition toward a complex information-processing hierarchy limited to the brain (Sloboda, 1985; Deutsch, 1999; Huron, 2006; Levitin, 2006). This is sometimes discussed in terms of discrete cognitive modules that have been naturally selected to perform specific tasks related to the survival of the species (Fodor, 1983; Pinker, 1997; Coltheart, 1999), leading some scholars to postulate 1:1 mappings between anatomical brain regions and musical functions (Peretz and Coltheart, 2003; cf. Altenmüller, 2001). While this research has indeed produced a number of important insights, it has arguably tended to downplay the role of the environmentally situated body in the development of musicality as a cognitive domain (see Clarke, 2005; Johnson, 2007).

In recent years, new perspectives have emerged that place more focus on the embodied, ecological, and dynamical dimensions of musical cognition (e.g., Borgo, 2005; Clarke, 2005, 2012; Reybrouck, 2005, 2013; Leman, 2007; Jones, 2009; Krueger, 2013; Maes et al., 2014; Moran, 2014; Laroche and Kaddouch, 2015; Godøy et al., 2016; Schiavio and van der Schyff, 2016; Schiavio et al., 2016; Lesaffre et al., 2017). Recent research has also tended to weaken the modular hypothesis by emphasizing the plastic and self-organizing properties of the (musical) brain (Jäncke et al., 2001; Pantev et al., 2001; Münte et al., 2002; Lappe et al., 2008; Large et al., 2016). The past two decades have also seen the development of a "biocultural" hypothesis for the origins and nature of the musical mind that looks beyond the traditional nature-culture dichotomy (Cross, 1999, 2003; Killin, 2013, 2016a,b, 2017; van der Schyff, 2013a,b; Tomlinson, 2015). This approach draws on a range of research in theoretical biology, neuroscience, embodied and ecological cognition, and dynamical systems theory (DST), positing a more integrated model that sees biological and cultural dimensions as aspects of the same evolving system. Here the origin of music is not understood within a strict adaptationist framework. Rather, it is explained as an emergent phenomenon involving cycles of (embodied) interactivity with the social and material environment.

Our aim in the present article is to contribute to the theoretical discussion supporting the biocultural hypothesis by considering it through the lenses of the enactive approach to cognition. This perspective first emerged in the work of Varela et al. (1991) and has been developed more recently across a range of contexts (Thompson, 2007; Stewart et al., 2010; Colombetti, 2014; Di Paolo et al., 2017). Most centrally, the enactive approach posits a deep continuity between mind and life, where cognitive processes are explored in terms of how self-organizing living systems enact relationships with the environment that are relevant to their survival and well-being. It highlights the embodied and ecologically situated nature of living agents, as well as the active role they play in their own developmental processes. Importantly, the enactive approach sees cognitive and evolutionary processes as driven by a range of interacting factors, including the socio-cultural forms of activity that characterize the lives of more complex creatures such as ourselves (Malafouris, 2008, 2013, 2015). We suggest, therefore, that it may help to extend the biocultural hypothesis in various ways.

We begin by providing a brief overview of some key positions in the field of evolutionary musicology, discussing how many tend to adhere to the "nature-or-culture" dichotomy mentioned above. We then outline the biocultural hypothesis, reviewing supporting research and theory in theoretical biology, neuroscience, and ecological and embodied cognition. Here we place a special focus on Tomlinson's (2015) approach as, for us, it represents the current state of the art in the field. While we are largely in agreement with his position, we suggest that future work could benefit from exploring a wider range of perspectives in embodied-ecological cognition. With this in mind, we then discuss the enactive approach and consider how it might enhance the biocultural perspective. More specifically, we suggest that the enactive view could offer theoretical support and refinement to Tomlinson's claim that the origins of the musical mind should be sought for in the embodied dynamics of coordinated action that occurred within the developing socio-material environments of our ancestors—and not first in terms of cognitive processes involving (quasi-linguistic) representational mental content. Following this, we consider how the recently emerged 4E approach—which sees cognition as *embodied*, *embedded*, *enactive*, and *extended*—aligns with the biocultural perspective, offering some tentative possibilities for how this framework might guide future research associated with the biocultural approach. To conclude we briefly consider the implications this perspective may have for thought and action in practical musical contexts (e.g., music education). Before we begin, we would also like to note that although the enactive approach is being explored across several disciplines (see Stewart et al., 2010), it has only recently been adopted in musical contexts (Borgo, 2005; Silverman, 2012; Krueger, 2013, 2014; Matyja and Schiavio, 2013; Elliott and Silverman, 2015; Loaiza, 2016; Schiavio et al., 2016). Therefore, this article may also contribute to the development of the enactive perspective for musical research and theory more generally.

## EVOLUTIONARY MUSICOLOGY AND THE DICHOTOMY OF ADAPTATION

An important point of discussion in evolutionary musicology concerns whether musicality can be considered as a bona fide adaptation, or if it is better understood as a product of culture (Huron, 2001; Davies, 2012; van der Schyff, 2013a; Lawson, 2014; Honing et al., 2015; Killin, 2016a, 2017). Some researchers (including Darwin, 1871) have drawn on comparisons with music-like behavior in other animals, suggesting an adaptive function for music in mate selection and territorial display in our prehistoric ancestors (see Miller, 2000). It has been argued, however, that although music-like behavior in non-human animals (e.g., bird song) may well be a product of natural selection, these traits are not *homologous* with human music making, but rather are *analogous* (Pinker, 1997; Hauser and McDermott, 2003). Because of this, it is claimed that comparative studies involving more phylogenetically distant species may not have great relevance for understanding the biological origins of human musicality (McDermott and Hauser, 2005; but see Fitch, 2006). Additionally, evidence of "musical" behaviors in our closest primate relatives is often understood to be sparse. For some scholars, this suggests there was no properly musical phenotype prior to modern humans in the hominin line (Huron, 2001; Justus and Hutsler, 2005; Patel, 2008).

Such arguments have been used to support claims that music should not be conceived of as an adaptation, but rather as a product of culture (e.g., Sperber, 1996; Pinker, 1997). Here it is posited that music is dependent on cognitive structures (e.g., modules) and abilities that evolved to support properly adaptive functions in our ancestors (e.g., language, auditory scene analysis, habitat selection, emotion, and motor control—for a discussion see Trainor, 2015). Perhaps the strongest version of this approach is found in Pinker (1997), who argues that music is an "invention" designed to "tickle" these naturally selected aspects of our cognitive and biological nature. Music itself, however, has no adaptive meaning: From an evolutionary point of view, it is the auditory equivalent of "cheesecake"—a cultural invention that is pleasurable, but biologically useless. In line with this, it is suggested that music might be a kind of exaptation—where the original (i.e., adapted) function of a trait becomes co-opted to serve other purposes[1] (Davies, 2012). Thus, as Sperber (1996) posits, music may be understood as "parasitic on a cognitive module the proper domain of which pre-existed music and had nothing to do with it" (p. 142).

[1]The term "exaptation" refers to changes in the function of a given physiological or behavioral trait in the process of the biological evolution of an organism. The classic example is bird feathers, which originally evolved for thermoregulation, but were later co-opted for mating-territorial display, catching insects, and then flight. The developmental systems approach discussed below complicates the causal relation of adaptations and exaptations. Here they stand not in a linear sequence, but rather in a cyclical relationship, where the new uses of an adaptation associated with the exaptation may lead to secondary adaptations and so on (see Gould and Vrba, 1982; Anderson, 2007). Referring to the relationship between adaptations and exaptations Tomlinson (2015) writes, "the first are not necessarily prior to the second, since behaviors originating as exaptations might alter selective pressures in ways leading to new adaptations" (p. 36).

By contrast, other researchers have suggested the existence of cognitive modules that appear to be specialized for musical functions. For example, Peretz (1993, 2006, 2012) research in acquired *amusia* has led her to (cautiously) posit an innate music-specific module for pitch processing, suggesting that music may be as "natural" as language (Peretz, 2006). Such claims are countered by Patel (2008), who argues that evidence indicating the existence of adapted music specific modules may in fact be explained by (ontogenetic) developmental processes, whereby cortical areas become specialized for certain functions through experience (e.g., via processes of "progressive modularization"; see Karmiloff-Smith, 1992). However, while Patel (2008, 2010) maintains that musicality in humans is not a "direct target" of natural selection, he also acknowledges the profound biological and social benefits associated with musical activity, claiming that music is a powerful "transformative technology of the mind" (Patel, 2008, p. 400–401). Here Patel discusses how musical experience may lead to long-lasting changes in brain structure and processing (e.g., though neuroendocrine effects). Interestingly, he also notes that the phenomenon of infant babbling, the anatomy of the human vocal tract, and the fixation of the FOXP2 gene, might be indicative of adaptations that originally supported both language and vocal music (Patel, 2008, p. 371–372). However, he suggests that because language appears to emerge more quickly and uniformly in humans, and because the lack of musical ability does not appear to entail significant biological costs, these factors are better understood to support the adaptive status of language. In brief, he posits that musical processing is a "by-product" of cognitive mechanisms selected for language and other forms of complex vocal learning (see also Patel, 2006, 2010, 2012).

These last claims are questioned by those who argue that they may reflect a rather narrow perspective on what musicality entails—e.g., the assumption that musical activity necessarily requires special forms of training, or that music is a pleasure product to be consumed at concerts or through recordings (for discussions see Small, 1999; Cross, 2003, 2010; van der Schyff, 2013a,b; Honing et al., 2015). With regard to this point, ethnomusicological and sociological research has revealed musical activity around the world to be central for human well-being—it is inextricable from work, play, social life, religion, ritual, politics, healing, and more (Blacking, 1973, 1995; Nettl, 1983, 2000; DeNora, 2000). Moreover, in many cultural environments music is highly improvisational in character, and the acquisition of musical skills begins in infancy and develops rapidly, often without the need for formal instruction (Blacking, 1973; Cross, 2003; Solis and Nettl, 2009). It has also been suggested that because certain physical and cognitive deficits need not hinder survival and well-being in modern Western society, certain "musical" impairments may go almost completely unnoticed (van der Schyff, 2013a). Likewise, music's relevance for human survival across evolutionary time has been considered in terms of its importance for bonding between infants and primary caregivers, and between members of social groups (Benzon, 2001; Tolbert, 2001; Dissanayake, 2010; Dunbar, 2012). Musical developmental processes appear to begin very early on in life (Parncutt, 2006) and researchers have demonstrated the universal

and seemingly intuitive way caregivers create musical (or music-like) environments for infants through prosodic speech and lullabies (Dissanayake, 2000; Trehub, 2003; Falk, 2004). Along these lines, Trevarthen (2002) has proposed that humans possess an in-born "communicative musicality" that serves the necessity for embodied inter-subjectivity in highly social beings such as ourselves (see also Malloch and Trevarthen, 2010).

In all, it is argued that the wide range of activities associated with the word "music" may have immediate and far-reaching implications for survival and socialization for many peoples of the world, as it may have had for our prehistoric ancestors (see Blacking, 1973; Mithen, 2005). And indeed, the archeological record shows evidence of musical activity (i.e., bone flutes) dating back at least 40,000 years (Higham et al., 2012; Morley, 2013). Such concerns drive the "musilanguage" theory put forward by Mithen (2005) and others (Brown, 2000; Lawson, 2014), where both music and language are understood to have developed from a "proto-musical ancestor" that evolved due to selective pressures favoring more complex forms of social behavior—e.g., enhanced types of communication associated with foraging and hunting, mate competition, increased periods of child rearing (soothing at a distance), and more complex forms of coordinated group activity (Dunbar, 1996, 2003, 2012; Cross, 1999, 2003; Falk, 2000, 2004; Balter, 2004; Bannan, 2012). Here it is also suggested that musical behavior may have contributed to the development of shared intentionality and Theory of Mind (ToM) in modern humans, which in turn permitted the rapid development of cultural evolution and the emergence of modern human cognition (Tomasello, 1999; Tomasello et al., 2005).

## THE BIOCULTURAL HYPOTHESIS

Thus far, we have offered only a brief outline of some of the main positions in the discussion over the status of music in human evolution. We would like to suggest, however, that although many important and well-reasoned accounts have emerged on both sides of the debate, the nature-or-culture perspective that appears to frame this discussion renders both sides somewhat problematic. On one hand, arguing that music is primarily a product of culture may tend to downplay its deep significance for human well-being, as well as the rather rapid and intuitive ways it develops in many cultural contexts. Indeed, as we have just considered, these manifold developmental and social factors are taken to be indicative of the biological relevance of music for the human animal. On the other hand, arguments for music as an adaptation (e.g., Mithen, 2005; Lawson, 2014) often tend to posit a singular adaptive status for what is in fact a complex phenomenon that spans a wide range of biological, social, and cultural dimensions (Tomlinson, 2015).

In line with such concerns, other scholars (Cross, 1999, 2001, 2003; Killin, 2013, 2016a; van der Schyff, 2013a,b; Currie and Killin, 2016) have offered alternative "biocultural" approaches to the nature and origins of human musicality—where the question of whether *either* biology *or* culture should account for deeply social and universal human activities that require complex cognitive functions (e.g., music) is replaced by a perspective that integrates the two. For example, Cross (1999) suggests that musicality is an emergent activity—or "cognitive capacity"—that arises from a more fundamental human proclivity to search for relevance and meaning in our interactions with the world. It is claimed that because of its "multiple potential meanings" and "floating intentionality" music provides a means by which social activity may be explored in a "risk free" environment, affording the development of competencies between different domains of embodied experience and the (co)creation of meaning and culture (Cross, 1999, 2003). Tomlinson (2015) develops similar insights, arguing that what we now refer to as "language" and "music" began with more basic forms of coordinated socio-cultural activity that incrementally developed into more sophisticated patterns of thought, activity, and communication (see also Morley, 2013). Moreover, such activities are understood to have transformed environmental niches over time (Sterelny, 2014; Killin, 2016a, 2017) and with them the behavioral possibilities (affordances) of the hominines who inhabited them through recursive cycles of feedback and feedforward effects.

In all, this orientation suggests a way through the traditional nature-or-culture dichotomy discussed above. In doing so, however, it necessarily draws on models of evolution and cognition that differ from those that have traditionally guided evolutionary musicology. In line with this, Tomlinson's (2015) approach develops Neo-Peircean perspectives in semiotics (e.g., Deacon, 1997, 2010, 2012), exploring how embodied and indexical forms of communication may in fact underpin our linguistic and musical abilities both in evolutionary and ontogenetic terms. As we discuss below, this is further supported by work in theoretical biology associated with developmental systems theory, studies of musical and social entrainment (rhythm and mimesis), and insights from ecological psychology and embodied cognition.

### Looking Beyond Adaptation

Tomlinson (2015) argues that although music-as-adaptation perspectives all reveal important aspects of why music is meaningful for the human animal, they are also problematic when they tend to assume a "unilateral explanation for a manifold phenomenon" (p. 33; see also Killin, 2016a). That is, because music takes on so many forms, involves such a wide range of behavior, and serves so many functions, it seems difficult to specify a single selective environment for it. And thus, these traits sit "uneasily side by side, their interrelation left unspecified" (p. 33). To be clear, this does not in any way negate the claims regarding the social and developmental meanings of music. These biologically relevant traits do exist, but they are just too numerous and complex to be properly described in terms of an adaptation (at least not in the orthodox sense of the term). Because of this, Tomlinson (2015) claims that we must be careful about how we frame evolutionary questions—and especially those regarding complex behaviors such as music and language—lest we fall into the reductive theorizing associated with "adaptationist fundamentalism." He thus argues that dwelling on the question of the adaptive status of music has had the effect of "focusing our sights too narrowly on

the question of natural selection alone—and usually a threadbare theorizing of it, at that" (p. 34).

With this in mind, the developmental systems approach to biological evolution posits a useful alternative perspective (see Oyama et al., 2001). In contrast to the one-directional schema that characterizes more traditional frameworks (where evolution is understood to involve *adaptation to* a given environment), developmental systems theory presents a more recursive and relational view, where organism and environment are understood as mutually influencing aspects of the same integrated system. Here evolutionary processes do not entail the adaptation of a species' phenotype to a fixed terrain, but rather "a dynamic interaction where other species and the non-living environment take part" (Tomlinson, 2015, p. 35). In other words, this approach explores the complex ways genes, organisms, and environmental factors—including behavior and (socio-cultural) experience—interact with each other in guiding the formation of phenotypes and the construction of environmental niches (Moore, 2003; Jablonka and Lamb, 2005; Richerson and Boyd, 2005; Malafouris, 2008, 2013, 2015; Laland et al., 2010; Sterelny, 2014). As such, it eschews the classic nature-nurture dichotomy, preferring instead to examine the interaction between organism and environment as a recursive or "dialectical" phenomenon (Lewontin et al., 1984; Pigliucci, 2001), where no single unit or mechanism is sufficient to explain all processes involved.

Importantly, the organism is understood here to play an active role in shaping the environment it *coevolves with*—its activities feedback into and alter the selective pressures of the environmental niche. This, in turn, affects the development of the organism, resulting in a co-evolutionary cycle that proceeds in an ongoing way. Socio-cultural developments add additional epicycles involving patterns of behavior that can sometimes hold stable over long periods of time (see **Figure 1**). These are passed on inter- and intra-generationally through embodied mimetic processes (more on this below; see also Sterelny, 2012). While such epicycles necessarily emerge from the coevolution cycle, they may, once established, develop into self-sustaining patterns of behavior that develop relatively independently. However, the effects of these cultural epicycles may feedforward into the broader coevolutionary system resulting in additional alterations to environmental conditions and shifts in biological configurations (e.g., gene expression and morphological changes—see Wrangham, 2009; Laland et al., 2010; Skinner et al., 2015; Killin, 2016a).

The making and use of tools is offered as a primary example of what such cultural epicycles might entail (Tomlinson, 2015). The archeological record contains many examples of bi-face stone hand axes that were made by our Paleolithic ancestors. These tools are remarkably consistent in their functional and aesthetic qualities, implying method and planning in their manufacture (Wynn, 1996, 2002). However, it is now thought that the production of these axes entailed a "bottom up" process based on the morphology and motor-possibilities of the body, unplanned emotional-mimetic social interaction, and the affordances of the environment (Gamble, 1999; Davidson, 2002). In other words, it is argued that the emergence of Paleolithic technologies did not involve abstract or representational forms of thought (e.g., a mental template, or "top down" thinking)—a capacity these early toolmakers did not possess (but see Killin, 2016b, 2017). Nor were they the result of genetically determined developmental programs. Rather, they are thought to have originated, developed, and stabilized primarily through the dynamic interaction between living systems and the material environments they inhabited and shaped (Ingold, 1999). It is suggested that such self-organizing forms of social-technological behavior provided the grounding from which more complex cultural activities like music emerged much later (Tomlinson, 2015). To better understand how this could be so, we now consider the mimetic nature of these pre-human social environments, and how this may give clues to the origins of music in coordinated rhythmic behavior.

## Mimesis, Entrainment, and the Origins of Music in Rhythm

In social animals, attention tends to be turned "outwards" toward the world and the activities of others (McGrath and Kelly, 1986). This entails the capacity to observe, understand, and emulate the actions of conspecifics. It is suggested that in our Paleolithic ancestors these mimetic processes allowed increasingly complex chains of actions to be passed on from one individual or generation to the next (Leroi-Gourhan, 1964/1993; Gamble, 1999; Ingold, 1999). This involved the enactment of culturally embedded "action loops" (see Donald, 2001; Tomlinson, 2015) that depended on a basic proclivity for forms of social *entrainment*.

The phenomenon of entrainment may be observed in many ways and over various timescales in both biological and non-biological contexts (de Landa, 1992; Clayton et al., 2005; Becker, 2011; Knight et al., 2017). Most fundamentally, it is understood in terms of the tendency for oscillating systems to synchronize with each other[2]. Accordingly, biological and social systems can be conceived of as dynamically interconnected systems of oscillating components (from metabolic cycles to life cycles, from single neuron firing to regional patterns of activity in the brain, from individual organisms to social groups and the broader biological and cognitive ecology; McGrath and Kelly, 1986; Oyama et al., 2001; Varela et al., 2001; Ward, 2003; Chemero, 2009). Importantly, the components of such systems influence each other in a non-linear or recursive way. As such, organism and environment are not separate domains, but rather aspects of "one non-decomposable system" that evolves over time (Chemero, 2009, p. 26). Moreover, the development of coupled systems is guided by local and global constraints that allow the system to maintain stability—to be resistant to perturbations, or to regain stability once a perturbation has occurred. This is, of course, crucial for living systems, which must maintain metabolic functioning within certain parameters if they are to survive.

Such self-organizing processes result in "emergent properties"—relationships, structures, and patterns of behavior

[2]A simple example of this is found in the way wall mounted pendulums mutually constrain one another, resulting in synchronization or "entrainment" over time (see Clark, 2001).

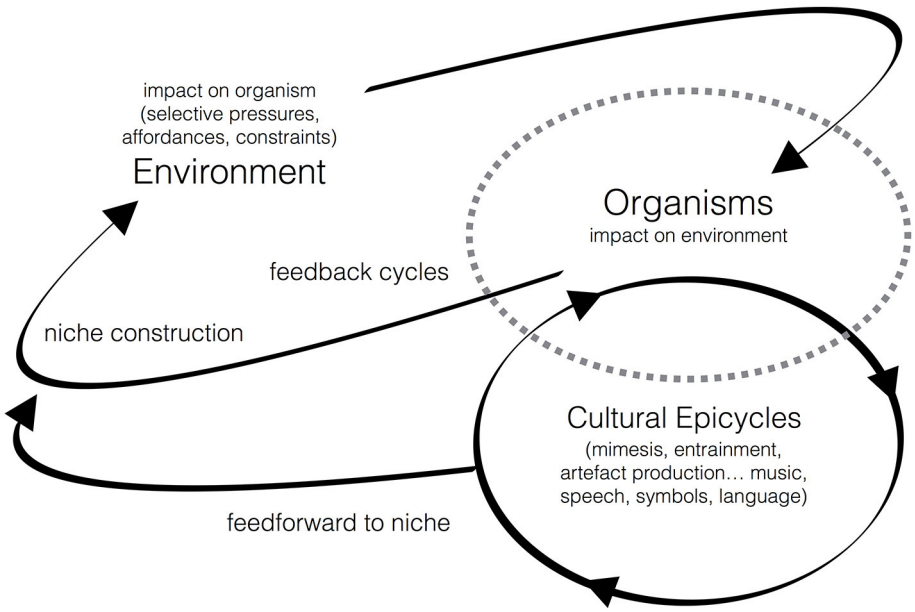



**FIGURE 1 |** Describes the cyclical process of biocultural coevolution (adapted with permission from Tomlinson, 2015, p. 46–47). Note that this depicts the most general level of description and does not show the more micro-level "cycles within cycles" that occur, for example, within the intra-organism milieu. These include the patterns of muscular, emotional-affective, neural, and metabolic activity that influence the expression of genes and gene groups over various timescales. This, in turn, helps to guide developmental processes and behavior that impacts the environmental niche.

that may remain consistent over long temporal periods, or that may be subject to transformation due to shifts in local and global constraints of the system. The mathematical techniques associated with DST have aided researchers in modeling such phenomena. Here patterns of convergence (stability) in the state of the system are contrasted with areas exhibiting entropy (instability; de Landa, 1992). This is often represented as a topographic "phase-space" that describes the possible states of a given system over time—periods of convergence in the trajectories of the system are represented as "basins of attraction" (Abraham and Shaw, 1985; Chemero, 2009). A "phase transition" occurs when new patterns of convergence arise (i.e., new attractor layouts). Researchers associated with developmental systems theory (above) use DST methods to model the evolutionary trajectories of coupled organism-environment systems, mapping dynamic patterns of stability and change as functions of constraint parameters (see Oyama et al., 2001).

DST is also used to examine how social animals bring their actions in line with those of other agents—and with other exogenous factors—by "dynamically attending" to the environment through sight, sound, movement, and touch (McGrath and Kelly, 1986; Large and Jones, 1999). This results in the enactment of coordinated forms of behavior that can occur both voluntarily and involuntarily. Emotional-affective aspects may also come into play here. For example, when a stable pattern is disrupted, entropy emerges in the system and a negative affect may result. The (living) system then self-organizes toward regaining stability, resulting in a positive effect. It is suggested that the action loops associated with Paleolithic toolmaking emerged from these forms of social entrainment—where dynamic couplings between various trajectories in the social environment led to increasingly stable patterns of behavior (basins of attraction) in the cultural epicycle. This permitted the mimetic transmission of cultural knowledge without the need for symbols, referentiality, or representation (see Tomlinson, 2015, p. 75).

Interestingly, the idea of dynamic attending has been explored empirically in the context of musical (i.e., metrical, rhythmic) entrainment (Large and Jones, 1999; Jones, 2009; Large et al., 2015). Tomlinson (2015) suggests that such dynamical models may help to reveal the distant origins of musical rhythm in the mimetic, emotional, and sonic-social environments jointly enacted by the coordinated (entrained) motor patterns of early toolmakers. This insight is supported by a range of current research into the evolution of rhythmic behavior (Fitch, 2012; Merchant and Bartolo, 2017; Ravignani et al., 2017). Indeed, evolutionary musicology has often tended to explore the origins of music in terms of its vocal dimensions (i.e., music as pitch/song production and its relationship to spoken language), and has thus had to wrestle with the issues associated with complex vocal learning, and its apparent absence in other primates. The focus on rhythm, however, has shown similarities between animal and human behavior (Fitch, 2010; Patel and Iversen, 2014; Merchant et al., 2015; Bannan, 2016; Iversen, 2016; Wilson and Cook, 2016). A large number of papers have also explored the deep relationship between rhythmic behavior and social cohesion in both human and non-human subjects (e.g., Large and Gray, 2015; Yu and Tomonaga, 2015; Tunçgenç and Cohen, 2016; Knight et al.,

2017). Additionally, recent studies by Ravignani et al. (2016a) have modeled the cultural evolution of rhythm in the lab. This research shows how, when presented with random percussive sounds, participants tend to develop structured and recurrent rhythms from such information, and that these patterns continue to develop through subsequent generations of participants who are asked to imitate the rhythms of previous generations. Interestingly, the rhythmic patterns that emerged in this study display six statistical universals found across different musical cultures and traditions. This aligns with the conception of cultural transmission based on mimesis and entrainment just discussed. It also implies that the enactment of musical (or music-like) behavior may not be traceable solely to the genome, but rather arises due to a more general propensity to structure acoustical experience in certain ways (see also Fitch, 2017).

Here it should be noted that the biocultural approach also develops a theory about the origins of vocal musicality, albeit one that is deeply connected to the rhythmic factors just described. This entails the development of a repertoire of "gesture-calls" similar to those found in modern primates and many mammalian species (grunts, pants-hoots, growls, howls, barking, and so on; see Tomlinson, 2015, p. 89–123). These do not involve the abstract, symbolic-representational, and combinatorial properties employed by modern languages. Rather, they are tightly coupled with the same mimetic, emotional, and embodied forms of communication that characterize pre-human tool-making. It is suggested that the vocal expressions associated with these gesture-calls reflected the sonic aspects (rhythmic and timbral) of these environments, the motor-patterns of production, as well as the gestural and social rhythms (e.g., turn taking, social entrainment) that developed within the cultural ecology. In line with this, studies show connections between rhythmic capacities and the development of vocal forms of communication, including language (Cummins and Port, 1996; Cummins, 2015; Bekius et al., 2016; Ravignani et al., 2016b). As an aside, it is also posited that the process of knapping may have resulted in specific forms of listening (Morley, 2013, p. 120), and that the resonant and sometimes tonal qualities of stones and flakes may have afforded music-like play with sound (Zubrow et al., 2001; Killin, 2016a,b)[3]. In brief, these rhythmic forms of behavior may have led to proto-musical and proto-linguistic forms of communication that arose simultaneously.

However, as Tomlinson (2015) notes, "half a million years ago there was no language or musicking" (p. 127). While many music-relevant anatomical features were in place by this period, there is no evidence that these hominins possessed the more complex forms of combinatorial thinking required for the hierarchical structuring of rhythm, timbre, and pitch associated with musical activity (i.e., the kind of thinking that is also needed to build tools specifically intended for musical use, such as bone flutes). Rather, it is posited that proto-musical and proto-linguistic communications were initially limited to deictic co-present interactions (in-the-moment face-to-face encounters that integrated gesture and a limited number of vocal utterances) that incrementally developed into more complex sequences of communicative behavior. Over time, this led to the enactment of increasingly sophisticated forms of joint action and social understanding (Dunbar, 1996, 2003; Knoblich and Sebanz, 2008; Sterelny, 2012). Such developments in the cultural loop fed forward into the coevolutionary cycle, allowing the environmental niche to be explored in new ways, affording previously unrecognized modes of engagement with it. This, in turn, altered selective pressures, leading to incremental phase transitions in the dynamics of the system, where previous constraints were weakened and new behavioral-cognitive phenotypes became possible. By the Upper Paleolithic period, the growing influence of the cultural epicycle favored an enhanced capacity to understand the actions and intentions of others and the related capacity to think "offline," "top down," or "at a distance" from immediate events (Bickerton, 1990, 2002; Carruthers and Smith, 1996; Tomasello, 1999). These developments allowed for the marshaling of material and social resources in new ways, leading to the creation of more complex artifacts (e.g., musical instruments), as well as more sophisticated types of cultural activity (e.g., ritual) and communication, including the hierarchical and combinatorial forms required for language and music as we know them today[4].

## Plastic Brains

The biocultural approach sees (musical) cognition as an emergent property of situated embodied activity within a developing socio-material environment. Because of this, it requires a rather different view of cognition than the information-processing model associated with an adapted (modular) brain (e.g., Fodor, 1983, 2001; Tooby and Cosmides, 1989, 1992; Pinker, 1997; Barrett and Kurzban, 2006). Indeed, if evolutionary processes do not involve adaption to a pre-given environment, but rather require the active participation of organisms in shaping the environments they *coevolve with*—where "selection" and "adaptation" are now understood in a contingent and dynamically cyclical context—then it seems reasonable to suggest that cognitive processes might not depend on genetically programmed responses or be reducible to a collection of fixed information-processing mechanisms in the brain. Rather, they might entail more plastic and perhaps non-representational characteristics that reflect the dynamic integration of brains, bodies, objects, and socio-cultural environments (for similar arguments see Malafouris, 2008, 2013, 2015).

In line with such concerns, scholars are questioning whether the notion of modularity continues to have much relevance for understanding the complexities of the human brain (e.g., Uttal, 2001; Doidge, 2007; Anderson, 2014). For example, it is suggested that brain regions that appear to consistently correlate with specific processes, such as Broca's area and syntax, represent vast

[3]Readers may be interested to consider the studies that examine the "musical" properties of stone artifacts and acoustical characteristics of Paleolithic environments (see Blake and Cross, 2008).

[4]This involves the integration of phonemes and words into grammatical structures and the development of a generative syntax that provides the "rules" for such processes—or, likewise the organization of discrete sets of sounds, tones, and pitches into rhythmic/formal hierarchies that could be consciously repeated or manipulated (e.g., melodies and drumming patterns).

areas of the cortex that may in fact develop multiple overlapping or interlacing networks, the manifold functions of which may appear evermore fine-grained and plastic as neural imaging technology becomes more refined (Hagoort, 2005; Poldrack, 2006; Tettamanti and Weniger, 2006; Grahn, 2012). In relation to this, recent research suggests the existence of "global systems" that function in a flexible and context-dependent manner (see Besson and Schön, 2012, p. 289–290). These do not work independently of any other information available to the brain and are thus non-modular (i.e., they are not discrete). Additionally, research into various levels of biological organization is showing that biological and cognitive processes develop in interaction with the environment—e.g., that epigenetic factors play a central role in the expression of genes, and that the formation of neural connections unfolds as a function of context (Sur and Leamey, 2001; Uttal, 2001; Van Orden et al., 2001; Lickliter and Honeycutt, 2003; Panksepp, 2009). In short, the idea that brain and behavior are best understood as linear systems decomposable into discrete modules and corresponding functions is being replaced by more plastic[5] and dynamically interactive perspectives. Such insights have contributed to the growing view that music cognition is the result of *non-modular* cognitive developmental processes that are driven by a more general attraction to coordinated forms of social behavior (Trehub, 2000; Trehub and Nakata, 2001-2002; Trehub and Hannon, 2006; see also Drake et al., 2000).

Because of this, recent decades have seen researchers turn to "connectionist" models to account for essential cognitive functions such as (musical) perception and learning (see Desain and Honing, 1991, 2003; Griffith and Todd, 1999; Clarke, 2005). Likewise, Tomlinson discusses the connectionist approach as a way of understanding how the embodied-ecological processes of mimesis and social entrainment contributed to the development of music and language. Put simply, the connectionist strategy does not rely on the idea of fixed modules, but rather on the fact that when simple devices (such as individual neurons) are massively interconnected in a distributed way such connections may change and grow through "experience"—when neurons tend to become active together, their connections are reinforced and vice versa (Hebb, 1949). Such connectivity is thought to result in the emergence of complex sub-systems of activity as well as global convergences that produce system wide properties. This is often modeled using DST and can also be understood in terms of the oscillatory dynamics mentioned above (see Chemero, 2009).

## Embodied Minds

While the connectionist approach was initially seen as an alternative to the computational orientation, more recent modeling has revealed the ability of complex connectionist networks to simulate syntactic, representational, and combinatorial cognitive processes (see Smolensky, 1990; Bechtel, 2008)—i.e., those required by the "adapted brain" hypothesis. Such developments are attractive for some researchers as they allow for the assumed computational-representational nature of cognition to remain while accommodating the growing evidence around brain plasticity and dynamism (Chalmers, 1990; Smolensky, 1990; Dennett, 1991; Clark, 1997; on compositionality see van Gelder, 1990). However, others maintain that because the brain's connectivity cannot be separated from its dynamic history of coupling with the body and the environment, living cognition is not best understood as strictly limited to in-the-brain computations and representational content (Varela et al., 1991; Thompson, 2007; Chemero, 2009; Hutto and Myin, 2012).

To better understand what this means for the biocultural approach to music's origins, it may be useful to consider Tomlinson's (2015, p. 129–139) reading of Cheney and Seyfarth's (2008) research into the social lives of baboons. As Tomlinson notes, observations of baboon vocal and gestural interactions lead Cheney and Seyfarth to suggest that the social behavior of these animals is indicative of an underlying hierarchical and syntactic-representational cognitive structure—one that is continuous with the Fodorian notion of "the language of thought" or "mentalese" (a process of non- or pre-conscious symbolic manipulation in the brain according to syntactic rules). This, they suggest, may reveal a deep evolutionary connection between linguistic processing and social intelligence—where linguistic-computational processes are thought to underpin social cognition even if no spoken or symbolic language is present (as with baboons and our pre-human ancestors; cf. Barrett, forthcoming). However, Cheney and Seyfarth also hint at another possibility, where a more plastic and dynamic connectionist framework comes into play. The idea here is that once a system learns to organize itself in various ways, the patterns it develops can be recognized by the system in association with various things and relationships and thus may be said to "represent" them[6]. For this reason, connectionist processes are sometimes thought to be "sub-symbolic" in that they provide a link between biological processes at lower levels and representational processes at higher ones (Varela et al., 1991, p. 100; Smolensky, 1988). In line with this, Cheney and Seyfarth (2008) suggest that as animals engage with their environments neural networks could be reinforced leading to multimodal forms of "distributed neural representation" (p. 241; see also Barsalou, 2005; Tomlinson, 2015, p. 133). As Tomlinson (2015) points out, this implies something less abstract and more concretely embodied and ecological:

> [A] *quite literal re-representing, a solidifying, affirming, salience-forming set of neural tautologies.* There is no reliance on abstracted social identities such as those humans conceive, on a mysterious language of mind that does the representing, or on baboon comprehension of causality, proposition, and predication. In their place are the accretion of intrabrain and interbrain networks and the responses they enable in face of situations that are both familiar and less so. Networks are, within sheer biological constraints, products of environmental affordances, forged through the repeated patterns of an organism's interaction with the socio material surroundings. [...] All the intricacy Cheney

[5]For studies on music and brain plasticity (see Large and Jones, 1999; Jäncke et al., 2001; Pantev et al., 2001; Schlaug, 2001; Münte et al., 2002; Gaser and Schlaug, 2003; Lappe et al., 2008). Additionally, clinical studies have demonstrated music's deep effects on the body as well as its capacity to transform or reorganize neural structures (e.g., Bunt, 1994; Standley, 1995; Nayak et al., 2000; Tomaino, 2009; Jovanov and Maxfield, 2011).

[6]See Toiviainen (2000) for a discussion of this in the context of music AI.

> and Seyfarth find in baboon sociality may well be explained [...] without recourse to anything like mentalese (p. 135–136; italics original).

Similarly, when Tomlinson (2015) refers to the mimetic nature of the developing proto-musical environments, he clarifies that the action loops associated with this may indeed be representational, but not in the sense of mental templates or propositions. Rather, following Donald (2001), Tomlinson comments that the notion of "representation" employed here may entail little more "than the rise to salience of an aspect of a hominin's environment—in this case an enacted sequence of physical gestures imprinting itself in neural networks that fire again when repeated. Or [...] a set of interconnected neural oscillations" (p. 73–74).

It is suggested that this revised conception of representation might be more conducive to understanding cognition across a wider range of developmental and phylogenetic contexts. The problem with applying the more traditional approach associated with computational psychology is that it tends to encourage a kind of "reverse engineering, retrospectively projecting human capacities onto earlier hominins or onto nonhuman species understood as proxies for our ancestors" (Tomlinson, 2015, p. 138). This critique resonates with the work of Barrett (2011), who discusses our tendency to construct highly anthropomorphic views of other life forms and how this can lead to false understandings—not only of their cognitive capacities, but also of the nature and origins of human minds. Similarly, it is argued that the traditional assumption that "cognition" necessarily involves some form of linguistic competence (syntax, propositional thought, symbolic representation, and other forms of abstract "mental gymnastics") has tended to overshadow the more fundamental embodied and emotional aspects of living meaning making in human cognition (Johnson, 2007). This extends to music, which over the past three decades has been examined with a special emphasis on its relationship to linguistic capacities in cognitive and evolutionary contexts (Patel, 2008; Rebuschat et al., 2012; van der Schyff, 2015).

Now, all of this is not meant to imply that research into the (cognitive and evolutionary) relationship between music and language should be abandoned. This is an important area of inquiry and should continue to be investigated. However, other developmental and socio-cultural factors are receiving growing attention from researchers. This includes accounts that explore the dynamic, ecological, and embodied nature of musical experience (e.g., Large and Jones, 1999; Reybrouck, 2005; Leman, 2007; Krueger, 2013; van der Schyff, 2015; Godøy et al., 2016). As we began to consider above, while music and language both involve hierarchical and combinatorial forms of thought, it may be that both emerge from more domain general capacities and proclivities related to the ways embodied-affective relationships are generated within socio-material environments (Johnson, 2007). For some scholars, this implies that the symbolic-representational and propositional forms of cognition associated with language may be derivative rather than primary (see Hutto and Myin, 2012, 2017). As such, the origins of cognition might not be found in brain-bound computations and symbolic representations, but rather in the self-organizing dynamics associated with biological development itself—in the cycles of action and perception that are directly linked to an organism's ongoing history of embodied engagement with its environment. This recalls the coevolution cycle discussed above, but it may also be considered in the context of ontogenesis—e.g., how infants enact meaningful realities through embodied and affective interactivity with their socio-material niche (see Bateson, 1975; Service, 1984; Dissanayake, 2000; Reddy et al., 2013).

Such insights are not lost on Tomlinson (2015), who highlights the continuity between the embodied activities of Paleolithic tool makers and cognition as such—where cognition might in fact be rooted in interactions with the environment that over time result in increasingly complex extensions of individual embodied minds into the broader cognitive ecology (e.g., via mimesis and social "rhythmic" entrainment). Here Tomlinson also entertains the possibility that the self-organizing (or "self-initiating" as he sometimes refers to it) nature of the activities discussed above might not need to be understood in representational terms at all. However, he does not go much further than this general suggestion. This is perhaps somewhat surprising as he does, here and there, draw on the notion of "affordances" and the field of ecological psychology it is associated with—an explicitly non-representational approach to cognition in its original version (Gibson, 1966, 1979; more on this shortly).

Once Tomlinson outlines the deeply embodied, ecological, and socially interactive precursors of musical behavior, he then turns to explain music cognition using generative (e.g., Lerdahl and Jackendoff, 1983) and prediction- or anticipation-based models (e.g., Huron, 2006) that focus on the (internal) processing of musical stimuli and the behavioral responses they lead to. These approaches are relevant to the discussion as they focus on the more abstract and combinatorial ways the modern human mind processes musical events. We would like to suggest, however, that future contributions might benefit by exploring a wider range of perspectives drawn from embodied cognitive science and related perspectives in music cognition. With this in mind, we now turn to discuss how insights associated with the enactive approach to cognition might help to support and advance many of the claims made by Tomlinson (2015) and the biocultural approach more generally.

## THE ENACTIVE PERSPECTIVE

The enactive approach to cognition was originally introduced by Varela et al. (1991) as a counter to the then dominant information-processing model of mind and the adaptationist approach to biological evolution[7]. Like the biocultural model, it develops the insights of developmental systems theory and DST, and is inspired by the work of Gibson (1966, 1979). Gibson's "ecological psychology" asks us to rethink the relationship

[7]It should be noted enactivism also has an antecedent in work by Bruner (1964) who coined the term.

between cognitive systems and their environment. As Chemero (2009) discusses, this can be understood in terms of three main tenets. The first posits that perception is direct (i.e., it is not mediated by representational mental content). The second argues that perception is not first and foremost for information gathering, but is for the guidance of action—for actively engaging with the world. Following from these, the third tenet claims that perception is of "affordances"—or the possibilities for action offered by the environment in relation to the corporeal complexity of the perceiving organism (e.g., a chair affords sitting for a child or an adult, but not for an infant or a fish; Gibson, 1979).

While sympathetic with the three core tenets of the Gibsonian approach, some scholars suggest that the conception of affordances associated with it is problematic when it implies that they are intrinsic features of the environment (e.g., Varela et al., 1991, p. 192–219; for a discussion see Chemero, 2009, p. 135–162). This, it is argued, does not give enough attention to the active role living creatures play in shaping the worlds they inhabit, leading "to a research strategy in which one attempts to build an ecological theory of perception entirely from the side of the environment. Such a research strategy ignores not only the structural unity of the animal but also the codetermination of animal and environment" (Varela et al., 1991, p. 204–205). In brief, the enactivist perspective posits a revised interpretation of affordances that more clearly integrates corporeal dimensions and the engaged perceptual activity of cognitive agents (Varela et al., 1991; see also Nöe, 2006; Chemero, 2009; Barrett, 2011; Schiavio, 2016). As we discuss next, this approach allows for a view of cognition that is not wholly driven by the environment—nor by internal representations—but rather by the embodied activity of living agents. As such, it may allow us to further develop the corporeal and ecological concerns that drive the biocultural model.

## Where There Is Life There Is Mind

One of the most central claims of the enactive perspective concerns the deep continuity between mind and life, where cognition is understood to originate in the self-organizing activity of living biological systems (Maturana and Varela, 1980, 1984; Varela et al., 1991; Thompson, 2007; Di Paolo et al., 2017). Most primarily, this involves the development and maintenance of a bounded metabolism (Jonas, 1966; Bourgine and Stewart, 2004; Thompson, 2007), but it also requires the (meta-metabolic) ability of the organism to move and interact with the environment in ways that are relevant to its survival (van Duijn et al., 2006; Egbert et al., 2010; Di Paolo et al., 2017; Barrett, forthcoming). Furthermore, because such fundamental life-processes occur under precarious conditions (Kyselo, 2014), they cannot be fully understood in an indifferent way. Rather, basic cognitive activity is characterized by a "primordial affectivity" that motivates relevant action (Colombetti, 2014). By this view, a living creature "makes sense" of the world through affectively motivated action-as-perception and, in the process, constructs a viable niche (Weber and Varela, 2002; Di Paolo, 2005; Reybrouck, 2005, 2013; Colombetti, 2010; Di Paolo et al., 2017). This involves the *enactment* of affordances—which are conceived of as emergent properties associated with the dynamic (evolutionary and ontogenetic) history of structural coupling between organisms and their environments[8] (Varela, 1988; Varela et al., 1991; Chemero, 2009; Barrett, 2011; Schiavio, 2016). Importantly, such basic sense-making processes do not involve the representational recovery of an external reality in the head (i.e., mental content). Rather, they are rooted in direct embodied engagement with the environment (Varela et al., 1991; Thompson, 2007)[9].

In brief, the enactive approach explores cognition in terms of the self-organizing and adaptive sense-making activities by which organisms enact survival-relevant relationships and possibilities for action (i.e., affordances) within a contingent milieu (Thompson, 2007). This constitutes the fundamental cognitive behavior of living embodied minds. Moreover, this perspective traces a continuity between the basic affectively motivated sense-making of simpler organisms and the richer manifestations of mind found in more complex biological forms (Di Paolo et al., 2017). In other words, where the meaningful actions of single-celled and other simple creatures are associated with factors related to nutrition and reproduction, more complicated creatures will engage in ever richer forms of sense making activity and thus exhibit a wider range of cognitive-emotional behaviors (Froese and Di Paolo, 2011). For social animals, this may include "participatory" forms of sense-making that involve the enactment of emotional-affective and empathic modes of communication between agents and social groups (mimesis), and that coincide with the development of shared repertoires of coordinated action (entrainment; see De Jaegher and Di Paolo, 2007; Di Paolo, 2009). With this in mind, we suggest that an enactive framework may provide a useful way of understanding human musical activities as continuous with, but not reducible to, the fundamental forms of self-organizing and emotionally driven action-as-perception that characterize living (participatory) sense-making more generally (van der Schyff, 2015; Loaiza, 2016; Schiavio and De Jaegher, 2017)[10]. As such,

[8]The symbiotic and co-emergent relationship between honeybees and flowers is an excellent example of this. Here autonomous organisms exist as environments to each other—the development of their phenotypes are inextricably enmeshed over evolutionary time (Varela et al., 1991; Hutto and Myin, 2012).

[9]This, of course, is not to say that the brain does not play an important role cognitive processes. However, from the enactive perspective, cognition is not limited to the brain—brain, body, and world are different aspects of an integrated cognitive system that functions in a non-linear way. Barrett (2011, p. 57–93) offers many examples that show how creatures with simple neural organizations are nevertheless able to engage in complex intelligent behaviours by using their bodies and environmental features as part of their cognitive systems (see also Brooks, 1991). In line with this, DST research into forms of problem solving and cognitive development associated with coordinated bodily activity have revealed that many of these processes can also be accurately described without necessarily having to recruit representational content (Thelen and Smith, 1994; Kelso, 1995; Chemero, 2009). Indeed, the DST equations employed to model such phenomena are neutral regarding representations. It is argued, therefore, that evoking representation may introduce unnecessary complications (see Chemero, 2009, p. 68–75).

[10]Among other things, this orientation has begun to offer insights into the ways the basic goal directed and self-organizing dynamics discussed above might be extended to living musical situations that are not life threatening in the literal sense, but that nevertheless require constant care and attention to maintain. For example, think of a performing string quartet. Each member must continuously adapt to the evolving musical environment, drawing on different

it appears to be well positioned to support and extend the biocultural model.

## Enactivism Meets the Biocultural Perspective

The enactive approach to cognition aligns with the biocultural model in several ways. Both draw on developmental systems theory and DST. And both embrace a circular and co-emergent view of organism and environment, as well as a deeply embodied perspective on cognition. Because the enactive approach traces cognition to the fundamental biological concerns shared by all forms of life, it may also help us avoid the anthropomorphizing tendencies noted above (e.g., imposing language-like capacities on non- or pre-human animals; but see De Jesus, 2015, 2016; Cummins and De Jesus, 2016), and thus better understand how cognitive capacities rooted in bodily action might ground the development of music and other cultural activities (Barrett, 2011; Tomlinson, 2015).

In connection with this, researchers drawing on enactivist theory are using DST models to examine bio-cognitive processes in terms of the non-linear couplings that occur between:

(i) the *body*—the development of muscular linkages and repertoires of corporeal articulation.
(ii) the *brain*—the emergence of patterned or recurrent (i.e., convergent) trajectories in neural activity.
(iii) the *environment*—the enactment of stable relationships and coordinated behavior within the socio-material ecology.

This approach is being explored across a range of areas (see Fogel and Thelen, 1987; Laible and Thompson, 2000; Hsu and Fogel, 2003; Camras and Witherington, 2005), including, for example, emotion research (Lewis and Granic, 2000; Colombetti, 2014), studies of social cognition and inter subjectivity (for a detailed discussion see Froese, forthcoming), and musical creativity (Walton et al., 2014, 2015). We suggest that similar approaches might be employed in conjunction with existing knowledge of early hominin anatomical and social structure, evidence from the archeological record, as well as comparative studies with other species and existing musical activities. This could also be developed alongside recent studies of how musical environments and behavior affect the expression of genes and gene groups, and how this might recursively influence behavioral and ecological factors (see Bittman et al., 2005, 2013; Schneck and Berger, 2006; Laland et al., 2010; Kanduri et al., 2015; Skinner et al., 2015).

Additionally, while recent theory associated with "radical enactivism" (Hutto and Myin, 2012) argues that so-called "basic minds" do not themselves possess any form of representational content, it also suggests that culture and language impose certain constraints that result in cognitive activities that may be understood as content bearing (this echoes the suggestion introduced above regarding the possible non-primary or "secondary" status of representational cognition; see Hutto and Myin, 2017). The explanatory advantages of this approach are currently a subject of debate. Nevertheless, the insights that arise from this discussion might shed new light on the cultural epicycles discussed above. As Tomlinson (2015) points out, although musical activity is not fundamentally symbolic or representational itself, it necessarily occurs and develops within cultural worlds of symbols and language. Put simply, the debate surrounding radical enactivism could offer new perspectives on how, over various developmental periods, cultural being might simultaneously constrain, and be driven by, the non-symbolic, social-affective, and embodied forms of cognition that characterize musical activity.

Another important possibility for how the enactive orientation might contribute to the biocultural approach involves the recently developed 4E framework, which sees cognition in terms of four overlapping dimensions—*embodied*, *embedded*, *enactive*, and *extended* (Menary, 2010a; Newen et al., 2017). The *embodied* dimension explores the central role the body plays in driving cognitive processes. This is captured, for example, in the description of the early Paleolithic tool making societies, where the reciprocal influences of sight, sound, and coordinated movement lead to the production of artifacts with specific characteristics. Such forms of embodied activity also formed the basis from which more complex forms of thought and communication emerged later. As we also considered, the biocultural model explores how such embodied factors arise in specific environments, leading to stable and recurrent patterns of activity where bodily, neural, and ecological trajectories converge. This highlights the *embedded* dimension, which concerns the ecological and socio-cultural factors that co-constitute situated cognitive activity. The biocultural model explores this in terms of the sonic, visual, tactile and emotional-mimetic nature of the niches enacted by our early ancestors, as well as the growing influence of the cultural epicycle on the cognitive ecology. The *enactive* dimension, as we have seen, concerns the self-organizing nature of living systems, and describes the active role organisms play in shaping the environments they inhabit. Such modes of activity (which are described as "sense-making") are explored over a range of timescales (brief encounters, ontogenesis, evolutionary development), closely aligning with the coevolutionary feedback cycle discussed above. As enactivists equate "sense-making" with "cognition" (Thompson, 2007; De Jaegher, 2013), it may be argued that mental life cannot be limited to the brains or bodies of organisms: It extends into the environments in which cognitive processes play out. In line with this, the *extended* dimension explores how many cognitive processes involve coupling with other agents (mimesis, social entrainment, participatory sense-making) or with non-biological objects or cultural artifacts (tools, notebooks, musical instruments; see Menary, 2010b; Malafouris, 2013, 2015). While Tomlinson (2015) makes no mention of enactivism or this 4E framework, he does, as we have seen, discuss how cognitive processes emerged and developed in our Paleolithic ancestors through

forms of embodied, emotional-affective/cognitive capacities to communicate, develop shared affordances, and maintain the musical ecology they co-create (this example is developed in detail by Salice et al., 2017; see also Krueger, 2014; Schiavio and Høffding, 2015). Similar studies by Walton et al. (2014, 2015) draw on enactive and dynamical systems theory to better understand the real-time dynamics of interacting musical agents in creative improvisational contexts.

embodied activity that was situated within a milieu that they actively shaped. He also argues that such activity necessarily involved the coordination of multiple agents and the "extension" of individual minds into the socio-material environment. We suggest, therefore, that a 4E approach might be useful in terms of organizing theoretical concepts and for framing and interpreting relevant empirical research.

The 4E framework is currently being developed by a handful of scholars in association with musical cognition (e.g., Krueger, 2014, 2016; Schiavio and Altenmüller, 2015; van der Schyff, 2017; Linson and Clarke, forthcoming). It is also explored in biological contexts by Barrett (2011, 2015a,b, forthcoming) as an alternative to the brain bound (and arguably anthropomorphizing) approach of traditional computationalism. Additionally, the 4E approach aligns with, and could be used to integrate, the corporeal, neural, and environmental levels of investigation associated with contemporary DST research in musical contexts. Therefore, it could help model how these factors contributed to the development of musical behavior in pre- and early human societies. Likewise, this approach might also have interesting implications for the laboratory modeling of cultural rhythmic transmission. As we began to discuss above, experiments by Ravignani et al. (2016a) examine how individuals trying to imitate random drumming sequences learn from each other in independent transmission chains—where the attempts of one participant become the training set for the next subject. This research aligns with the biocultural and enactive perspectives when it suggests that cultural development is not the product of genetic programming, but is guided by more general dynamical processes and constraints that allow for a range of possibilities. A 4E approach might develop the parameters of such studies to include the manipulation of social environmental (i.e., embedded + extended) factors—possibly exploring how groups of participants (rather than chains of individual drummers) collaboratively make sense of their sonic environments and develop rhythmic patterns in real time, and how the shared environments that result are transmitted and developed (enacted) by the following cohort. Additionally, it might be interesting to introduce different instruments and methods of sound making it to the environment to see how this affects the results. Lastly, a 4E approach could also include the analysis of video and audio recordings to better understand the relationship between the (embodied) motor, sonic, and socio-material factors involved in the enactment of "rhythmic cultures"[11]. If it is indeed the case that it is joint bodily action that drove cognitive and cultural processes in our ancestors, then it would be interesting to see how drumming movements shape shared learning environments, and how they develop into new more structured ones (more efficient and easier to imitate) as the rhythmic patterns are transmitted.

[11]A relevant example of approaches involving the integration of video and audio documentation, and DST/4E analysis, may be found in the recent work by Walton et al. (2014, 2015) that examines perceptions of creativity in interacting musical improvisers (see also Borgo, 2005; Laroche and Kaddouch, 2015). Note that these studies also include a phenomenological dimension that incorporates first-person accounts of the participants.

## CONCLUSION

We have offered here only a few tentative possibilities for how the enactive and 4E orientation might extend the biocultural approach to the origins and nature of human musicality. We hope that the ideas we have discussed here will inspire future work that explores this relationship more fully. Along these lines, readers may be interested to consider recent work by Malafouris (2008, 2013, 2015), who develops enactive and 4E principles to better understand how brains, bodies, and objects interact to form cognitive ecologies. Malafouris expands the idea of neural plasticity discussed above to include the domain of objects, tools, and culture. In doing so he posits a notion of "metaplasticity" that demands an "historical ontology" of different forms of material engagement (Malafouris, 2013, 2015). This is considered at the intersection of neuroscience, archeology, 4E cognition, and approaches to biological evolution that are closely aligned with developmental systems theory. In many ways, Malafouris' perspective sums up the interests and aspirations of the biocultural approach. He writes,

> I propose to accept the fact that human cognitive and emotional states literally comprise elements in their surrounding material environment. Our attention, therefore, should shift from the distinction of "mind" and "matter" or "in" and "out," toward developing common relational ways of thinking about the complex interactions among brain, body, and world. If we succeed, traditional ways of doing cognitive science should change, and the change will stretch far beyond the context of cognitive archaeology and human evolution (Malafouris, 2015, p. 366).

With this in mind, we would like to close by briefly mentioning some ontological and ethical implications an enactive-biocultural model might have for practical areas like music education. If music is neither a pleasure technology, nor the result of some strict adaptationist process—but rather a biocultural phenomenon rooted in the dynamics of joint action—then the ways we approach it in practice (e.g., music education, musicology, performance, music therapy, and so on) should reflect this fundamental existential reality. In other words, this approach opens a perspective on what it means to be and become musical that is no longer based in prescriptive developmental processes, adapted cognitive modules, and correspondence to pre-given stimuli (e.g., music as the reproduction of a score; see Small, 1999). Instead, it highlights the plastic, creative, situated, participatory, improvisational, embodied, empathic, and world-making nature of human musicality. It may therefore offer support to a growing number of theorists who argue that we have tended to rely on disembodied, depersonalized, and highly "technicist" approaches to musical learning (Regelski, 2002, 2016; Borgo, 2007; Elliott and Silverman, 2015), and that this orientation has reduced the ontological status of music students, teachers, listeners, and performers to mere responders, consumers, and reproducers (van der Schyff et al., 2016). Although this cannot be explored in detail here, it is an example of how alternative perspectives on the evolution

and nature of human (musical) cognition could inspire new ways of thinking in practical areas. In all, then, we hope that the biocultural and enactive approaches will continue to be developed in musical contexts to gain richer understandings of the origins and meaning of musicality for the human animal.

## AUTHOR CONTRIBUTIONS

DvdS developed the main body of text. AS provided suggestions and comments that were implemented in the final version.

## FUNDING

DvdS is supported by the Social Sciences and Humanities Research Council of Canada.

## ACKNOWLEDGMENTS

We thank Luca Barlassina and Richard Parncutt for inspiring discussions concerning this topic. We also thank the Action Editor, Aleksey Nikolsky, and the reviewers for their helpful comments.

## REFERENCES

Abraham, R., and Shaw, C. (1985). *Dynamics: The Geometry of Behaviour, Vol. 3.* Santa-Cruz, CA: Aerial Press.

Altenmüller, E. (2001). How many music centers are in the brain? *Ann. N. Y. Acad. Sci.* 930, 273–280. doi: 10.1111/j.1749-6632.2001.tb05738.x

Anderson, M. (2007). Massive redeployment, exaptation, and the functional integration of cognitive operations. *Synthese* 159, 329–345. doi: 10.1007/s11229-007-9233-2

Anderson, M. L. (2014). *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA; London: MIT Press.

Balter, M. (2004). Seeking the key to music. *Science* 306, 1120–1122. doi: 10.1126/science.306.5699.1120

Bannan, N. (ed.). (2012). *Music, Language, and Human Evolution*. Oxford, UK: Oxford University Press. doi: 10.1093/acprof:osobl/9780199227341.001.0001

Bannan, N. (2016). Darwin, music and evolution: new insights from family correspondence on The Descent of Man. *Mus. Sci.* 21, 3–25. doi: 10.1177/1029864916631794

Barrett, H. C., and Kurzban, R. (2006). Modularity in cognition: framing the debate. *Psychol. Rev.* 113, 628–647. doi: 10.1037/0033-295X.113.3.628

Barrett, L. (2011). *Beyond the Brain: How Body and Environment Shape Animal and Human Minds*. Princeton, NJ: Princeton University Press.

Barrett, L. (2015a). A better kind of continuity. *South. J. Philos. Spind. Suppl.* 53, 28–49. doi: 10.1111/sjp.12123

Barrett, L. (2015b). Why brains aren't computers, why behaviourism isn't Satanism and why dolphins aren't aquatic apes. *Behav. Anal.* 39, 9–23. doi: 10.1007/s40614-015-0047-0

Barrett, L. (forthcoming). "The evolution of cognition: a 4E perspective," in *The Oxford Handbook of 4e Cognition*, eds A. Newen, L. De Bruin, and S. Gallagher (New York, NY: Oxford University Press).

Barsalou, L. W. (2005). Continuity of the conceptual system across species. *Trends Cogn. Sci. (Regul. Ed).* 9, 309–311. doi: 10.1016/j.tics.2005.05.003

Bateson, M. C. (1975). "Mother infant exchanges: the epigenesis of conversational interaction," in *Annals of the New York Academy of Sciences: Developmental Psycholinguistics and Developmental Disorders*, Vol. 263, eds D. Aronson and R. W. Reiber (New York, NY: New York Academy of Sciences), 101–113.

Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York, NY: Routledge.

Becker, J. (2011). "Rhythmic entrainment and evolution," in *Music, Science and the Rhythmic Brain: Cultural and Clinical Implications*, eds J. Berger and G. Turow (London: Routledge), 49–72.

Bekius, A., Cope, T. E., and Grube, M. (2016). The beat to read: a cross-lingual link between rhythmic regularity perception and reading skill. *Front. Hum. Neurosci.* 10:425. doi: 10.3389/fnhum.2016.00425

Benzon, W. L. (2001). *Beethoven's Anvil: Music in Mind and Culture*. New York, NY: Basic Books.

Besson, M., and Schön, D. (2012). "What remains of modularity?" in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. Hawkins, and I. Cross (Oxford: Oxford University Press), 283–291.

Bickerton, D. (1990). *Language and Species*. Chicago, IL: University of Chicago Press.

Bickerton, D. (2002). "Foraging versus social intelligence in the evolution of protolanguage," in *The Transition to Language,* ed A. Wray (Oxford: Oxford University Press), 207–225.

Bittman, B., Berk, L., Shannon, M., Sharaf, M., Westengard, J., Guegler, K. J., et al. (2005). Recreational music-making modulates the human stress response: a preliminary individualized gene expression strategy. *Med. Sci. Monit.* 11, 31–40.

Bittman, B., Croft, D. T. Jr., Brinker, J., van Laar, R., Vernalis, M. N., and Ellsworth, D. L. (2013). Recreational music-making alters gene expression pathways in patients with coronary heart disease. *Med. Sci. Monit.* 19, 139–147. doi: 10.12659/MSM.883807

Blacking, J. (1973). *How Musical is Man?* London: Faber.

Blacking, J. (1995). *Music, Culture, and Experience*. Chicago, IL: University of Chicago Press.

Blake, E. C., and Cross, I. (2008). "Flint tools as portable sound-producing objects in the upper palaeolithic context: an experimental study," in *Experiencing Archaeology by Experiment*, eds P. Cunningham, J. Heeb, and R. Paardekooper (Oxford: Oxbow Books), 1–19.

Borgo, D. (2005). *Sync or Swarm: Improvising Music in a Complex Age*. New York, NY: Continuum.

Borgo, D. (2007). Free jazz in the classroom: an ecological approach to music education. *Jazz Perspect.* 1, 61–88. doi: 10.1080/17494060601061030

Bourgine, P., and Stewart, J. (2004). Autopoiesis and cognition. *Artif. Life,* 10, 327–345. doi: 10.1162/1064546041255557

Brooks, R. (1991). Intelligence without representation. *Artif. Intell.* 47, 139–159. doi: 10.1016/0004-3702(91)90053-M

Brown, S. (2000). "The 'musilanguage' model of human evolution," in *The Origins of Music,* eds N.L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 271–300.

Bruner, J. S. (1964). The course of cognitive growth. *Amer. Psychol.* 19, 1–15. doi: 10.1037/h0044160

Bunt, L. (1994). *Music Therapy: An Art Beyond Words.* London: Routledge. doi: 10.4324/9780203359716

Camras, L. A., and Witherington, D. C. (2005). Dynamical system approaches in emotional development. *Dev. Rev.* 25, 328–350. doi: 10.1016/j.dr.2005.10.002

Carruthers, P., and Smith, P. K. (1996). *Theories of Theories of Mind*. Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511597985

Chalmers, D. (1990). "Why Fodor and Pylyshyn were wrong: the simplest refutation," in *Proceedings of the 12th Annual Conference of the Cognitive Science Society* (Hillsdale, NJ: Lawrence Erlbaum).

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: The MIT Press.

Cheney, D., and Seyfarth, R. (2008). *Baboon Metaphysics: The Evolution of a Social Mind*. Chicago, IL: University of Chicago Press.

Clark, A. (1997). *Being there: Putting Brain Body and World Together*. Cambridge, MA: MIT Press.

Clark, A. (2001). *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford: Oxford University Press.

Clarke, E. F. (2005). *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. New York, NY: Oxford University Press.

Clarke, E. F. (2012). "What's going on: music, psychology and ecological theory," in *The Cultural Study of Music*, eds M. Clayton, T. Herbert, and R. Middleton (London: Routledge), 333–342.

Clayton, M., Sager, R., and Will, U. (2005). In time with the music: the concept of entrainment and its significance for ethnomusicology. *Eur. Meet. Ethnomusicol.* 11, 3–142.

Colombetti, G. (2010). "Enaction, sense-making and emotion," in *Enaction: Toward a New Paradigm for Cognitive Science*, eds J. R. Stewart, O. Gapenne, and E. A. Di Paolo (Cambridge, MA: MIT Press), 145–164.

Colombetti, G. (2014). *The Feeling Body: Affective Science Meets the Enactive Mind*. Cambridge, MA: The MIT Press.

Coltheart, M. (1999). Modularity and cognition. *Trends Cogn. Sci.* 3, 115–120. doi: 10.1016/S1364-6613(99)01289-9

Cross, I. (1999). "Is music the most important thing we ever did? Music, development and evolution," in *Music, Mind and Science,* ed Suk Won Yi (Seoul: Seoul National University Press), 10–39. Available online at: http://www-personal.mus.cam.ac.uk/~ic108/MMS/index.html

Cross, I. (2001). Music, mind and evolution. *Psychol. Music* 29, 95–102. doi: 10.1177/0305735601291007

Cross, I. (2003). "Music and biocultural evolution," in *The Cultural Study of Music*, eds M. Clayton, T. Herbert, and R. Middleton (London: Routledge), 17–27.

Cross, I. (2010). "The evolutionary basis of meaning in music: some neurological and neuroscientific implicationsm," in *The Neurology of Music,* ed F. Clifford (London: Imperial College Press), 1–15.

Cummins, F. (2015). "Rhythm and speech," in *Blackwell Handbooks in Linguistics. The Handbook of Speech Production, 1st Edn.,* ed M. A. Redford (Hoboken, NY: Wiley), 158–177. doi: 10.1002/9781118584156.ch8

Cummins, F., and De Jesus, P. (2016). The loneliness of the enactive cell: towards a bio-enactive framework. *Adapt. Behav.* 24, 149–159. doi: 10.1177/1059712316644965

Cummins, F., and Port, R. F. (1996). "Rhythmic commonalities between hand gestures and speech," in *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (London: Lawrence Erlbaum Associates), 415–419.

Currie, A., and Killin, A. (2016). Musical pluralism and the science of music. *Eur. J. Philos. Sci.* 6, 9–30. doi: 10.1007/s13194-015-0123-z

Darwin, C. R. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: John Murray.

Davidson, I. (2002). "The 'finished artefact fallacy': Acheulean handaxes and language origins," in *Transitions to Language,* ed A. Wray (Oxford: Oxford University Press), 180–203.

Davies, S. (2012). *The Artful Species: Aesthetics, Art, and Evolution*. Oxford: Oxford University Press.

De Jaegher, H. (2013). Embodiment and sense-making in autism. *Front. Integr. Neurosci.* 7:15. doi: 10.3389/fnint.2013.00015

De Jaegher, H., and Di Paolo, E. A. (2007). Participatory sense-making: an enactive approach to social cognition. *Phenomenol. Cogn. Sci.* 6, 485–507. doi: 10.1007/s11097-007-9076-9

De Jesus, P. (2015). Autopoietic enactivism, phenomenology and the deep continuity between life and mind. *Phenomenol. Cogn. Sci.* 15, 265–289. doi: 10.1007/s11097-015-9414-2

De Jesus, P. (2016). Making sense of (autopoietic) enactive embodiment: a gentle appraisal. *Phainomena* 25, 33–56.

de Landa, M. (1992). "Non organic life," in *Zone 6: Incorporations,* eds J. Crary and S. Kwinter (New York, NY: Urzone), 129–167.

Deacon, T. W. (1997). *The Symbolic Species: The Co-evolution of Language and the Human Brain*. London: Penguin.

Deacon, T. W. (2010). *On the Human: Rethinking the Natural Selection of Human Language*. Available online at: https://nationalhumanitiescenter.org/on-the-human/2010/02/on-the-human-rethinking-the-natural-selection-of-human-language/

Deacon, T. W. (2012). *Incomplete Nature: How Mind Emerged from Matter*. New York, NY: Norton.

Dennett, D. (1991). *Consciousness Explained*. Boston, MA: Little, Brown.

DeNora, T. (2000). *Music in Everyday Life*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511489433

Desain, P., and Honing, H. (1991). "Quantization of musical time: a connectionist approach," in *Music and Connectionism,* eds P. M. Todd and G. Loy (Cambridge, MA: MIT Press), 150–167.

Desain, P., and Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception* 32, 341–365. doi: 10.1068/p3370

Deutsch, D. (1999). "The processing of pitch combinations," in *The Psychology of Music* (New York, NY: Academic Press), 349–411.

Di Paolo, E. A. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenol. Cogn. Sci.* 4, 97–125. doi: 10.1007/s11097-005-9002-y

Di Paolo, E. A. (2009). Extended life. *Topoi* 28, 9–21. doi: 10.1007/s11245-008-9042-3

Di Paolo, E. A., Buhrmann, T., and Barandiaran, X. E. (2017). *Sensorimotor Life: An Enactive Proposal*. New York, NY: Oxford University Press.

Dissanayake, E. (2000). "Antecedents of the temporal arts in early mother-infant interaction," in *The Origins of Music,* eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 389–410.

Dissanayake, E. (2010). "Root, leaf, blossom, or bole: concerning the origin and adaptive function of music," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C.Trevarthen (Oxford: Oxford University Press), 17–30.

Doidge, N. (2007). *The Brain that Changes Itself: Stories of Personal Triumph from the Frontiers of Brain Science*. New York, NY: Penguin.

Donald, M. (2001). *A Mind So Rare: The Evolution of Human Consciousness*. New York, NY: Norton.

Drake, C., Jones, M. R., and Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition* 77, 251–288. doi: 10.1016/S0010-0277(00)00106-2

Dunbar, R. (1996). *Grooming, Gossip and the Evolution of Language*. Cambridge, MA: Harvard University Press.

Dunbar, R. (2003). "The origin and subsequent evolution of language," in *Language Evolution,* eds M.H. Christiansen and S. Kirby (Oxford, UK: Oxford University Press). 219–234.

Dunbar, R. (2012). "On the evolutionary function of song and dance," in *Music, Language, and Human Evolution*, ed N. Bannan (Oxford, UK: Oxford University Press), 201–214.

Egbert, M., Barandiaran, X., and Di Paolo, E. (2010). A minimal model of metabolism-based chemotaxis. *PLoS Comput. Biol.* 6:e1001004. doi: 10.1371/journal.pcbi.1001004

Elliott, D. J., and Silverman, M. (2015). *Music Matters: A Philosophy of Music Education, 2nd Edn*. New York, NY: Oxford University Press.

Falk, D. (2000). "Hominid brain evolution and the origins of music," in *The Origins of Music,* eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 197–216.

Falk, D. (2004). Prelinguistic evolution in early hominins: whence motherese? *Behav. Brain Sci.* 27, 450–491. doi: 10.1017/S0140525X04000111

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Fitch, W. T. (2012). "The biology and evolution of rhythm: unraveling a paradox," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohmeier, J. A. Hawkins, and I. Cross (Oxford: Oxford University Press), 73–95.

Fitch, W. T. (2017). Cultural evolution: lab-cultured musical universals. *Nat. Hum. Behav.* 1, 1–2. doi: 10.1038/s41562-016-0018

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (2001). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

Fogel, A., and Thelen, E. (1987). Development of early expressive and communicative action: reinterpreting the evidence from a dynamical systems perspective. *Dev. Psychol.* 23, 747–761. doi: 10.1037/0012-1649.23.6.747

Froese, T. (forthcoming). "Searching for the conditions of genuine intersubjectivity: from agent-based models to perceptual crossing experiments," in *The Oxford Handbook of 4E Cognition,* eds A. Newen, L. de Bruin, and S. Gallagher (New York, NY: Oxford University Press).

Froese, T., and Di Paolo, E. A. (2011). The enactive approach: theoretical sketches from cell to society. *Pragmat. Cogn.* 19, 1–36. doi: 10.1075/pc.19.1.01fro

Gamble, C. (1999). *The Palaeolithic Societies of Europe*. Cambridge, UK: Cambridge University Press.

Gaser, C., and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *J. Neurosci.* 23, 240–245.

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton-Mifflin.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton-Mifflin.

Godøy, R. I., Song, M., Nymoen, K., Haugen, M. R., and Jensenius, A. R. (2016). Exploring sound-motion similarity in musical experience. *J. New Music Res.* 45, 210–222. doi: 10.1080/09298215.2016.1184689

Gould, S. J., and Vrba, E. S. (1982). Exaptation: a missing term in the science of form. *Paleobiology* 8, 4–15. doi: 10.1017/S0094837300004310

Grahn, J. (2012). "Advances in neuroimaging techniques: implications for the shared syntactic resource hypothesis," in *Language and Music as Cognitive Systems,* eds P. Rebuschat, M. Rohrmeier, J. Hawkins, and I. Cross (Oxford: Oxford, University Press), 235–241.

Griffith, N., and Todd, P. (eds.). (1999). *Musical Networks: Parallel Distributed Perception and Performance*. Cambridge, MA: MIT Press.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends Cogn. Sci. (Regul. Ed).,* 9, 416–423. doi: 10.1016/j.tics.2005.07.004

Hauser, M., and McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nat. Neurosci.* 6, 663–668. doi: 10.1038/nn1080

Hebb, D. O. (1949). *The Organization of Behavior*. New York, NY: Wiley & Sons.

Higham, T., Basell, L., Jacobi, R., Wood, R., Ramsey, C. B., and Conard, N. J. (2012). Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: the radiocarbon chronology of Geißenklösterle'. *J. Hum. Evol.* 62, 664–676. doi: 10.1016/j.jhevol.2012.03.003

Honing, H., ten Cate, C., Peretz, I., and Trehub, S. (2015). Without it no music: cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140088. doi: 10.1098/rstb.2014.0088

Hsu, H.-C., and Fogel, A. (2003). Stability and transitions in mother-infant face-to- face communication during the first 6 months: a microhistorical approach. *Dev. Psychol.* 39, 1061–1082. doi: 10.1037/0012-1649.39.6.1061

Huron, D. (2001). Is music an evolutionary adaptation? *Ann. N.Y. Acad. Sci.* 930, 43–61. doi: 10.1111/j.1749-6632.2001.tb05724.x

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Hutto, D., and Myin, E. (2012). *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge, MA: The MIT Press.

Hutto, D., and Myin, E. (2017). *Evolving Enactivism: Basic Minds Meet Content*. Cambridge MA: The MIT Press.

Ingold, T. (1999). Tools for the hand, language for the face: an appreciation of Leroi-Gourhan's Gesture and Speech. *Stud. History Philos. Sci. C* 30, 411–445. doi: 10.1016/S1369-8486(99)00022-9

Iversen, J. R. (2016). "In the beginning was the beat: evolutionary origins of musical rhythm in humans," in *The Cambridge Companion to Percussion,* ed R. Hartenberger (Cambridge: Cambridge University Press), 281–295.

Jablonka, E., and Lamb, M. (2005). *Evolution in Four Dimensions*. Cambridge MA: MIT Press.

Jäncke, L., Gaab, N., Wüstenberg, T., Scheich, H., and Heinze, H. J. (2001). Short-term functional plasticity in the human auditory cortex: an fMRI study. *Brain Res.* 12, 479–485 doi: 10.1016/S0926-6410(01)00092-1

Johnson, M. (2007). *The Meaning of the Body: Aesthetics of Human Understanding*. Chicago, IL: University of Chicago Press.

Jonas, H. (1966). *The Phenomenon of Life*. Chicago: IL: University of Chicago Press.

Jones, M. R. (2009). "Musical time," in *Oxford Handbook of Music Psychology, 1st Edn.*, eds S. Hallam, I. Cross, and M. Thaut (Oxford: OxfordUniversity Press), 81–92.

Jovanov, E., and Maxfield, M. C. (2011). "Entraining the brain and body," in *Music, Science and the Rhythmic Brain: Cultural and Clinical Implications,* eds J. Berger and G. Turow (London: Routledge), 31–48.

Justus, T., and Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: assessing innateness and domain-specificity. *Music Percept.* 23, 1–27. doi: 10.1525/mp.2005.23.1.1

Kanduri, C., Kuusi, T., Ahvenainen, M., Philips, A. K., Lähdesmäki, H., and Järvelä, I. (2015). The effect of music performance on the transcriptome of professional musicians. *Sci. Rep.* 5:9506. doi: 10.1038/srep09506

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Approach to Cognitive Science*. Cambridge, MA: MIT Press.

Kelso, S. (1995). *Dynamic Patterns.* Cambridge, MA: MIT Press.

Killin, A. (2013). The arts and human nature: evolutionary aesthetics and the evolutionary staus of art behaviors. *Biol. Philos.* 28, 703–718. doi: 10.1007/s10539-013-9371-5

Killin, A. (2016a). Rethinking music's status as adaptation versus technology: a niche construction perspective. *Ethnomusicol. Forum* 25, 210–233. doi: 10.1080/17411912.2016.1159141

Killin, A. (2016b). Musicality and the evolution of mind, mimesis, and entrainment. *Biol. Philos.* 31, 421–434. doi: 10.1007/s10539-016-9519-1

Killin, A. (2017). Plio-pleistocene foundations of Hominin musicality: coevolution of cogition, sociality, and music. *Biol. Theor.* doi: 10.1007/s13752-017-0274-6. [Epub ahead of print].

Knight, S., Spiro, N., and Cross, I. (2017). Look, listen and learn: exploring effects of passive entrainment on social judgements of observed others. *Psychol. Music* 45, 99–115. doi: 10.1177/0305735616648008

Knoblich, G., and Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 2021–2031. doi: 10.1098/rstb.2008.0006

Krueger, J. (2013). "Empathy, enaction, and shared musical experience: evidence from infant cognition," in *The Emotional Power of Music: Multidisciplinary Perspectives on Musical Expression, Arousal, and Social Control*, eds T. Cochrane, B. Fantini, and K. Scherer (Oxford: Oxford University Press), 177–196.

Krueger, J. (2014). Affordances and the musically extended mind. *Front. Psychol.* 4:1003. doi: 10.3389/fpsyg.2013.01003

Krueger, J. (2016). "Musical worlds and the extended mind," in *Proceedings of A Body of Knowledge - Embodied Cognition and the Arts conference CTSA UCI*. Available online at: https://ore.exeter.ac.uk/repository/handle/10871/28061

Kyselo, M. (2014). The body social: an enactive approach to the self. *Front. Psychol.* 5:986. doi: 10.3389/fpsyg.2014.00986

Laible, D. J., and Thompson, R. A. (2000). "Attachment and self-organization," in *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development*, eds M. D. Lewis and I. Granic (Cambridge: Cambridge University Press), 298–323.

Laland, K. N., Odling-Smee, J., and Myles, S. (2010). How culture shaped the human genome: bringing genetics and the human sciences together. *Nat. Rev. Genet.* 11, 137–148. doi: 10.1038/nrg2734

Lappe, C., Herholz, S. C., Trainor, L. J., and Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *J. Neurosci.* 28, 9632–9639. doi: 10.1523/JNEUROSCI.2254-08.2008

Large, E. W., and Gray, P. M. (2015). Spontaneous tempo and rhythmic entrainment in a Bonobo (*Pan Paniscus*). *J. Comp. Psychol.* 129, 317–328. doi: 10.1037/com0000011

Large, E. W., and Jones, M. R. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119–159. doi: 10.1037/0033-295X.106.1.119

Large, E. W., Herrera, J. A., and Velasco, M. J. (2015). Neural networks for beat perception in musical rhythm. *Front. Syst. Neurosci.* 9:159. doi: 10.3389/fnsys.2015.00159

Large, E. W., Kim, J. C., Flaig, N., Bharucha, J., and Krumhansl, C. L. (2016). A neurodynamic account of musical tonality. *Music Percept.* 33, 319–331. doi: 10.1525/mp.2016.33.3.319

Laroche, J., and Kaddouch, I. (2015). Spontaneous preferences and core tastes: embodied musical personality and dynamics of interaction in a pedagogical method of improvisation. *Front. Psychol.* 6:522. doi: 10.3389/fpsyg.2015.00522

Lawson, F. R. S. (2014). Is music an adaptation or a technology? Ethnomusicological perspectives from the analysis of Chinese Shuochang. *Ethnomusicol. Forum* 23, 3–26. doi: 10.1080/17411912.2013.875786

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.

Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: The MIT Press.

Leroi-Gourhan, A. (1964/1993). *Gesture and Speech* (Transl. by A. Bostock Berger, Intro. by R. White). Cambridge, MA; London: MIT Press.

Lesaffre, M., Leman, M., and Maes, P. J. (eds.). (2017). *The Routledge Companion to Embodied Music Interaction*. New York, NY: Routledge.

Levitin, D. (2006). *This is Your Brain on Music: The Science of a Human Obsession*. New York, NY: Penguin.

Lewis, M. D., and Granic, I. (eds.). (2000). *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development*. Cambridge: Cambridge University Press.

Lewontin, R., Rose, S., and Kamin, L. (1984). *Not in Our Genes: Biology, Ideology and Human Nature*. New York, NY: Pantheon Books.

Lickliter, R., and Honeycutt, H. (2003). Developmental dynamics: toward a biologically plausible evolutionary psychology. *Psychol. Bull.* 129, 819–835. doi: 10.1037/0033-2909.129.6.819

Linson, A., and Clarke, E. F. (forthcoming). "Distributed cognition, ecological theory, and group improvisation," in *Distributed Creativity: Collaboration and Improvisation in Contemporary Music*, eds E. F. Clarke and M. Doffman (New York, NY: Oxford University Press).

Loaiza, J. M. (2016). Musicking, embodiment and the participatory enaction of music: outline and key points. *Conn. Sci.* 28, 410–422. doi: 10.1080/09540091.2016.1236366

Maes, P.-J., Leman, M., Palmer, C., and Wanderly, M. (2014). Action-based effects on music perception. *Front. Psychol.* 4:1008. doi: 10.3389/fpsyg.2013.01008

Malafouris, L. (2008). Beads for a plastic mind: the 'blind man's stick (BMS) hypothesis and the active nature of material culture. *Cambridge Archaeol. J.* 18, 401–414. doi: 10.1017/S0959774308000449

Malafouris, L. (2013). *How Things Shape the Mind: A Theory of Material Engagement*. Cambridge, MA: The MIT Press.

Malafouris, L. (2015). Metaplasticity and the primacy of material engagement. *Time Mind* 8, 351–371 doi: 10.1080/1751696X.2015.1111564

Malloch, S., and Trevarthen, C. (eds.). (2010). *Communicative Musicality: Exploring the Basis of Human Companionship*. Oxford: Oxford University Press.

Maturana, H., and Varela, F. (1984). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston, MA: New Science Library.

Maturana, H., and Varela, F. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Boston, MA: Reidel.

Matyja, J., and Schiavio, A. (2013). Enactive music cognition. Background and research themes. *Construct. Found.* 8, 351–357.

McDermott, J., and Hauser, M. D. (2005). The origins of music: innateness, development, and evolution. *Music Percept.* 23, 29–59. doi: 10.1525/mp.2005.23.1.29

McGrath, J. E., and Kelly, J. R. (1986). *Time and Human Interaction: Toward a Social Psychology of Time*. New York, NY: Guilford Press.

Menary, R. A. (2010a). Introduction to the special issue on 4E cognition. *Phenomenol. Cogn. Sci.* 9, 459–463. doi: 10.1007/s11097-010-9187-6

Menary, R. A. (ed.). (2010b). *The Extended Mind*. Cambridge, MA: The MIT Press.

Merchant, H., and Bartolo, R. (2017). Primate beta oscillations and rhythmic behaviors. *J. Neural Transm.* 124, 20–31 doi: 10.1007/s00702-017-1716-9

Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., and Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. B* 370:20140093. doi: 10.1098/rstb.2014.0093

Miller, G. F. (2000). "Evolution of human music through sexual selection," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: The MIT Press), 329–360.

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld & Nicholson.

Moore, D. (2003). *The Dependent Gene: The Fallacy of "Nature vs. Nurture."* New York, NY: Henry Holt.

Moran, N. (2014). Social implications arise in embodied music cognition research which can counter musicological "individualism." *Front. Psychol.* 5:676. doi: 10.3389/fpsyg.2014.00676

Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archeology, and the Origins of Musicality*. Oxford: Oxford University Press.

Münte, T. F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* 3, 473–478. doi: 10.1038/nrn843

Nayak, S., Wheller. B. L., Shiflett, S. C., and Agostinelli, S. (2000). Effect of music therapy on mood and social interaction among individuals with acute traumatic brain injury and stroke. *Rehabil. Psychol.* 45, 274–283. doi: 10.1037/0090-5550.45.3.274

Nettl, B. (1983). *The Study of Ethnomusicology: Twenty-nine Issues and Concepts*. Urbana, IL: University of Illinois Press.

Nettl, B. (2000). "An ethnomusicologist contemplates universals in musical sound and musical culture," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 463–472.

Newen, A., de Bruin, L., and Gallagher, S. (2017). *Oxford Handbook of 4E Cognition*. Oxford: Oxford University Press.

Nöe, A. (2006). *Action in Perception*. Cambridge, MA: The MIT Press.

Oyama, S., Griffiths, P., and Gray, R. (2001). *Cycles of Contingency: Developmental Systems and Evolution*. Cambridge, MA: MIT Press.

Panksepp, J. (2009). The emotional antecedents to the evolution of music and language. *Music. Sci.* Special Issue 2009–2010, 229–259. doi: 10.1177/1029864909013002111

Pantev, C., Engelien, A., Candia, V., and Elbert, T. (2001). Representational cortex in musicians: plastic alterations in response to musical practice. *Ann. N. Y. Acad. Sci.* 930, 300–314. doi: 10.1111/j.1749-6632.2001.tb05740.x

Parncutt, R. (2006). "Prenatal development," in *The Child as Musician: A Handbook of Musical Development*, ed G. E. McPherson (Oxford: Oxford University Press), 1–31.

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Percept.* 24, 99–104. doi: 10.1525/mp.2006.24.1.99

Patel, A. D. (2008). *Music, Language and the Brain.* New York, NY: Oxford University Press.

Patel, A. D. (2010). "Music, biological evolution, and the brain," in *Emerging Disciplines*, ed M. Bailar (Houston, TX: Rice University Press), 91–144.

Patel, A. D. (2012). "Language, music, and the brain: a resource-sharing framework," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. Hawkins, and I. Cross (Oxford: Oxford University Press), 204–223.

Patel, A. D., and Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front. Syst. Neurosci.* 8:57. doi: 10.3389/fnsys.2014.00057

Peretz, I. (1993). Audio atonalia for melodies. *Cogn. Neuropsychol.* 10, 21–56. doi: 10.1080/02643299308253455

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition* 100, 1–32. doi: 10.1016/j.cognition.2005.11.004

Peretz, I. (2012). "Music, language and modularity in action," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. Hawkins, and I. Cross (Oxford: Oxford University Press) 254–268.

Peretz, I., and Coltheart, M. (2003). Modularity of music processing. *Nat. Neurosci.* 6, 688–691. doi: 10.1038/nn1083

Pigliucci, M. (2001). *Phenotypic Plasticity: Beyond Nature and Nurture*. Baltimore, MD: Johns Hopkins University Press.

Pinker, S. (1997). *How the Mind Works.* New York, NY: Norton.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neruroimaging data? *Trends Cogn. Sci.* 10, 59–63. doi: 10.1016/j.tics.2005.12.004

Ravignani, A., Delgado, T., and Kirby, S. (2016a). Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* 1:0007. doi:10.1038/s41562-016-0007

Ravignani, A., Fitch, W. T., Hanke, F. D., Heinrich, T., Hurgitsch, B., Kotz, S. A., et al. (2016b). What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Front. Neurosci.* 10:274. doi: 10.3389/fnins.2016.00274

Ravignani, A., Honing, H., and Kotz, S. A. (2017). Editorial: The evolution of rhythm cognition: timing in music and speech. *Front. Hum. Neurosci.* 11:303. doi: 10.3389/fnhum.2017.00303

Rebuschat, P., Rohrmeier, M., Hawkins, J., and Cross, I. (2012). *Language and Music as Cognitive Systems*. Oxford: Oxford University Press.

Reddy, V., Markova, G., and Wallot, S. (2013). Anticipatory adjustments to being picked up in infancy. *PLoS ONE* 8:e65289. doi: 10.1371/journal.pone.0065289

Regelski, T. A. (2002). On "methodolatry" and music teaching as critical and reflective praxis. *Philos. Music Educ. Rev.* 10, 102–123. doi: 10.2979/PME.2002.10.2.102

Regelski, T. A. (2016). Music, music education, and institutional ideology: a praxial philosophy of music sociality. *Action Criticism Theory Music Educ.* 15, 10–45.

Reybrouck, M. (2005). "Body, mind and music: musical semantics between experiential cognition and cognitive economy," in *Trans: Transcultural Music Review,* Vol. 9. Available online at: http://www.sibetrans.com/trans/articulo/180/body-mind-and-music-musical-semantics-between-experiential-cognition-and-cognitive-economy

Reybrouck, M. (2013). From sound to music: an evolutionary approach to musical semantics. *Biosemiotics* 6, 585–606. doi: 10.1007/s12304-013-9192-6

Richerson, P. J., and Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago, IL: University of Chicago Press.

Salice, A., Høffding, S., and Gallagher, S. (2017). Putting plural self-awareness into practice: the phenomenology of expert musicianship. *Topoi* doi: 10.1007/s11245-017-9451-2. [Epub ahead of print].

Schiavio, A. (2016). Enactive affordances and the interplay of biological and phenomenological subjectivity. *Construct. Found.* 11, 315–317.

Schiavio, A., and Altenmüller, E. (2015). Exploring music-based rehabilitation for Parkinsonism through embodied cognitive science. *Front. Neurol.* 6:217. doi: 10.3389/fneur.2015.00217

Schiavio, A., and De Jaegher, H. (2017). "Participatory sense-making in joint musical practices," in *The Routledge Companion to Embodied Music Interaction,* eds M. Lesaffre, M. Leman, and P. J. Maes (New York, NY; London: Routledge), 31–39.

Schiavio, A., and Høffding, S. (2015). Playing together without communicating? A pre-reflective and enactive account of joint musical performance. *Music. Sci.* 19, 366–388. doi: 10.1177/1029864915593333

Schiavio, A., and van der Schyff, D. (2016). Beyond musical qualia: reflecting on the concept of experience. *Psychomusicology* 26, 366–378. doi: 10.1037/pmu0000165

Schiavio, A., van der Schyff, D., Cespedes-Guevara, J., and Reybrouck, M. (2016). Enacting musical emotions. sense-making, dynamic systems, and the embodied mind. *Phenomenol. Cogn. Sci.* doi: 10.1007/s11097-016-9477-8. [Epub ahead of print].

Schlaug, G. (2001). The brain of musicians. A model for functional and structural adaptation. *Ann. N.Y. Acad. Sci.* 930, 81–99. doi: 10.1111/j.1749-6632.2001.tb05739.x

Schneck, D. J., and Berger, D. S. (2006). *The Music Effect: Music Physiology and Clinical Applications*. London: Jessica Kingsley Publishers.

Service, V. (1984). "Maternal styles and communicative development," in *Language Development,* eds A. Lock and E. Fisher (London: Croom Elm), 132–140.

Silverman, M. (2012). Virtue ethics, care ethics, and "The good life of teaching". *Action Critic. Theory Music Educ.* 11, 96–122.

Skinner, M. M., Stephens, N. B., Tsegai, Z. J., Foote, A. C., Nguyen, N. H., Gross, T., et al. (2015). Human-like hand use in *Australopithecus africanus*. *Science* 347, 395–399. doi: 10.1126/science.1261735

Sloboda, J. A. (1985). *The Musical Mind: The Cognitive Psychology of Music*. Oxford: Clarendon.

Small, C. (1999). *Musicking: The Meaning of Performing and Listening*. Middletown, CT: Wesleyan University Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behav. Brain Sci.* 11, 1–74. doi: 10.1017/S0140525X00052432

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artif. Intell.* 46, 159–216. doi: 10.1016/0004-3702(90)90007-M

Solis, G., and Nettl, B. (2009). *Musical Improvisation: Art, Education, and Society*. Chicago, IL: University of Illinois Press.

Sperber, D. (1996). *Explaining Culture*. Oxford: Blackwell.

Standley, J. (1995). "Music as a therapeutic intervention in medical and dental treatment: research and applications," in *The Art and Science of Music Therapy: A Handbook,* eds T. Wigram, B. Saperston, and R. West (Amsterdam: Harwood Academic Publishers), 3–22.

Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. Cambridge, MA: MIT Press.

Sterelny, K. (2014). "Constructing the cooperative niche," in *Entangled Life: History, Philosophy and Theory of the Life Sciences*, Vol. 4, eds G. Barker, E. Desjardins, and T. Pearce (New York, NY: Springer), 261–279. doi: 10.1007/978-94-007-7067-6_13

Stewart, J., Gapenne, O., and Di Paolo, E. A. (eds.). (2010). *Enaction: Toward a New Paradigm for Cognitive Science*. Cambridge: The MIT Press.

Sur, M., and Leamey, C. A. (2001). Development and plasticity of cortical areas and networks. *Na. Rev. Neurosci.* 2, 251–262. doi: 10.1038/35067562

Tettamanti, M., and Weniger, D.,(2006). Broca's area: a supramodal hierarchical processor? *Cortex* 42, 491–494. doi: 10.1016/S0010-9452(08)70384-8

Thelen, E., and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Thompson, E. (2007). *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Toiviainen, P. (2000). "Symbolic AI versus connectionism in music research," in *Readings in Music and Artificial Intelligence,* ed E. R. Miranda (London: Routledge), 47–68.

Tolbert, E. (2001). Music and meaning: an evolutionary story. *Psychol. Music* 29, 84–94. doi: 10.1177/0305735601291006

Tomaino, C. M. (2009). "Clinical applications of music therapy in neurologic rehabilitation," in *Music that Works: Contributions of Biology, Neurophysiology, Psychology, Sociology, Medicine and Musicology*, eds R. Haas and V. Brandes (Vienna: Springer), 211–220.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691. doi: 10.1017/S0140525X05000129

Tomlinson, G. (2015). *A Million Years of Music: The Emergence of Human Modernity*. New York, NY: Zone Books.

Tooby, J., and Cosmides, L. (1989). Evolutionary psychology and the evolution of culture, part I. *Ethol. Sociobiol.* 10, 29–49. doi: 10.1016/0162-3095(89)90012-5

Tooby, J., and Cosmides, L. (1992). "The psychological foundation of culture," in *The Adapted Mind*, eds Barkow, L. Cosmides, and J. Tooby (Oxford: Oxford, University Press), 19–136.

Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philos. Trans. R. Soc. B* 370:2014.0089. doi: 10.1098/rstb.2014.0089

Trehub, S. E. (2000). "Human processing predispositions and musical universals," in *The Origins of Music,* eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 427–448.

Trehub, S. E. (2003). "Musical predispositions in infancy: an update," in *The Cognitive Neuroscience of Music*, eds I. Peretz and R. Zatorre (New York, NY: Oxford, University Press), 3–20.

Trehub, S. E., and Hannon, E. E. (2006). Infant music perception: domain-general or domain-specific mechanisms? *Cognition* 100, 73–99. doi: 10.1016/j.cognition.2005.11.006

Trehub, S. E., and Nakata, T. (2001-2002). Emotion and music in infancy. *Music. Sci.* 6, 37–61. doi: 10.1177/10298649020050S103

Trevarthen, C. (2002). "Origins of musical identity: evidence from infancy for musical social awareness," in *Musical Identities*, eds R. A. R. MacDonald, D. J. Hargreaves, and D. Miell (Oxford: Oxford, U. P), 21–38.

Tunçgenç B., and Cohen, E. (2016). Movement synchrony forges social bonds across group divides. *Front. Psychol.* 7:782. doi: 10.3389/fpsyg.2016.00782

Uttal, W. (2001). *The New Phrenology*. Cambridge, MA: MIT Press.

van der Schyff, D. (2013a). Music, culture and the evolution of the human mind: looking beyond dichotomies. *Hellenic J. Music Educ. Cult.* 4, 5–15.

van der Schyff, D. (2013b). *Music, Meaning, and the Embodied Mind: Toward an Enactive Approach to Music Cognition*. Unpublished MA thesis, University of Sheffield.

van der Schyff, D. (2015). Music as a manifestation of life: exploring enactivism and the 'eastern perspective' for music education. *Front. Psychol.* 6:345. doi: 10.3389/fpsyg.2015.00345

van der Schyff, D. (2017). Refining the model for emotion research: a 4E perspective. *Construct. Found.* 12, 227–229.

van der Schyff, D., Schiavio, A., and Elliott, D. J. (2016). Critical ontology for an enactive music pedagogy. *Action Critic. Theory Music Educ.* 15, 81–121. doi: 10.22176/act15.5.81

van Duijn et al. 2006 van Duijn, M., Keijzer, F., and Franken, D. (2006). Principles of minimal cognition: casting cognition as sensorimotor coordination. *Adapt. Behav.* 14, 157–170. doi: 10.1177/105971230601400207

van Gelder, T. (1990). Compositionality: a connectionist variation on a classical theme. *Cogn. Sci.* 14, 355–384. doi: 10.1207/s15516709cog1403_2

Van Orden, G. C., Pennington, B. F., and Stone, G. O. (2001). What do double disassociations prove? Modularity yields a degenerating research program. *Cogn. Sci.* 25, 111–117. doi: 10.1207/s15516709cog2501_5

Varela, F. (1988). "Structural coupling and the origin of meaning in a simple cellular automata," in *The Semiotics of Cellular Communications in the Immune System*, eds E. Secarez, F. Celada, N.A. Mitchinson, and T. Tada (New York, NY: Springer-Verlag), 151–161.

Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 4, 229–239. doi: 10.1038/35067550

Varela, F., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge MA: MIT Press.

Walton, A., Richardson, M. J., and Chemero, A. (2014). Self-organization and semiosis in jazz improvisation. *Int. J. Signs Semiot. Syst.* 3, 12–25. doi: 10.4018/IJSSS.2014070102

Walton, A., Richardson, M. J., Langland-Hassan, P., and Chemero, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Front. Psychol.* 1:313. doi: 10.3389/fpsyg.2015.00313

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* 12, 553–559. doi: 10.1016/j.tics.2003.10.012

Weber, A., and Varela, F. J. (2002). Life after Kant: natural purposes and the autopoietic foundations of biological individuality. *Phenomenol. Cogn. Sci.* 1, 97–125. doi: 10.1023/A:1020368120174

Wilson, M., and Cook, P. F. (2016). Rhythmic entrainment: why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychon. Bull. Rev.* 23, 1647–1659. doi: 10.3758/s13423-016-1013-x

Wrangham, R. (2009). *Catching Fire: How Cooking Made Us Human*. London: Profile.

Wynn, T. G. (1996). "The evolution of tools and symbolic behavior," in *Handbook of Human In Handbook of Human Symbolic Evolution*, eds A. Lock and C. R. Peters (Oxford: Clarendon Press), 263–287.

Wynn, T. G. (2002). Archaeology and cognitive evolution. *Behav. Brain Sci.* 25, 389–438. doi: 10.1017/S0140525X02000079

Yu, L., and Tomonaga, M. (2015). Interactional synchrony in chimpanzees: examination through a finger-tapping experiment. *Sci. Rep.* 5:10218. doi: 10.1038/srep10218

Zubrow, E., Cross, I., and Cowan, F. (2001). Musical behaviour and the archaeology of the mind. *Archaeol. Polona* 39, 111–126.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 van der Schyff and Schiavio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Neuroscience

HYPOTHESIS AND THEORY
published: 06 October 2017
doi: 10.3389/fnins.2017.00542

# The Role of the Baldwin Effect in the Evolution of Human Musicality

*Piotr Podlipniak**

*Institute of Musicology, Adam Mickiewicz University in Poznań, Poznań, Poland*

From the biological perspective human musicality is the term referred to as a set of abilities which enable the recognition and production of music. Since music is a complex phenomenon which consists of features that represent different stages of the evolution of human auditory abilities, the question concerning the evolutionary origin of music must focus mainly on music specific properties and their possible biological function or functions. What usually differentiates music from other forms of human sound expressions is a syntactically organized structure based on pitch classes and rhythmic units measured in reference to musical pulse. This structure is an auditory (not acoustical) phenomenon, meaning that it is a human-specific interpretation of sounds achieved thanks to certain characteristics of the nervous system. There is historical and cross-cultural diversity of this structure which indicates that learning is an important part of the development of human musicality. However, the fact that there is no culture without music, the syntax of which is implicitly learned and easily recognizable, suggests that human musicality may be an adaptive phenomenon. If the use of syntactically organized structure as a communicative phenomenon were adaptive it would be only in circumstances in which this structure is recognizable by more than one individual. Therefore, there is a problem to explain the adaptive value of an ability to recognize a syntactically organized structure that appeared accidentally as the result of mutation or recombination in an environment without a syntactically organized structure. The possible solution could be explained by the Baldwin effect in which a culturally invented trait is transformed into an instinctive trait by the means of natural selection. It is proposed that in the beginning musical structure was invented and learned thanks to neural plasticity. Because structurally organized music appeared adaptive (phenotypic adaptation) e.g., as a tool of social consolidation, our predecessors started to spend a lot of time and energy on music. In such circumstances, accidentally one individual was born with the genetically controlled development of new neural circuitry which allowed him or her to learn music faster and with less energy use.

**Keywords: Baldwin effect, human musicality, cortico-subcortical loops, pitch structure, musical rhythm**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky,*
*Braavo! Enterprises, United States*

***Reviewed by:***
*Nicholas Bannan,*
*University of Western Australia, Australia*
*Edward Hagen,*
*Washington State University Vancouver, United States*

***Correspondence:**
*Piotr Podlipniak*
*podlip@poczta.onet.pl*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *24 May 2017*
***Accepted:*** *19 September 2017*
***Published:*** *06 October 2017*

***Citation:***
*Podlipniak P (2017) The Role of the Baldwin Effect in the Evolution of Human Musicality. Front. Neurosci. 11:542. doi: 10.3389/fnins.2017.00542*

## INTRODUCTION

Human musicality can be understood as a set of abilities which enable people to recognize and produce music (Fitch, 2015). Although the worldwide diversity of music reveals the cultural flexibility of *Homo sapiens* in musical behavior, the fact that people in all known cultures sing (Nettl, 2000) and can recognize music without explicit learning (Tillmann, 2005) suggests that

musicality is a part of human biological endowment (Blacking, 1973; Fitch, 2006, 2015). In fact, a variety of theories have been proposed which try to explain the possible adaptive value of music (Cross and Morley, 2008). Many of the possible biological values of music have been put forward such as increasing sexual attractiveness (Darwin, 1871; Miller G. F., 2000), facilitating mother-infant bonds (Dissanayake, 2008), enhancing group consolidation (Roederer, 1984; Storr, 1992; Harvey, 2017), reducing cognitive dissonance (Perlovsky, 2010, 2012), and alerting outsiders about a group cohesiveness (Hagen and Bryant, 2003; Hagen and Hammerstein, 2009), to name only the most popular ideas. However, music is a complex communicative phenomenon (Zimmermann et al., 2013) which is composed of many features. Some of these features are shared with other communicative phenomena, which raises the question about music specificity and in consequence, about the biological character of human musicality. For example, the manipulation of sound intensity, stress and tempo is an important part of all human songs but also of speech. Moreover, a similar use of these features is observed among the vocal expressions of many mammalian species (Zimmermann et al., 2013) including our closest relative—the chimpanzee (*Pan troglodytes*) (Slocombe et al., 2009). The patterns of continuously modulated sounds used as a tool of expression and induction of emotions in other individuals are called "expressive dynamics" (Merker, 2003) or "affective prosody" (Zimmermann et al., 2013) and seem to be evolutionarily more ancient than any species of hominins. Of course, music is also usually composed of such elements which are not shared with other forms of human sound expressions. The interpretation of musical stimuli in terms of pitch classes and temporal isochrony are at least two musical elements which are absent in speech (Fitch, 2013a) and other human vocalizations. Although pitch and the duration of vowels can serve as discreet phonological units in tonal and time sensitive languages respectively (Remijsen and Gilley, 2008; Wong et al., 2012; Remijsen, 2014) both pitch and vowel length in speech lack hierarchical ordering based on mental reference points. Such a hierarchy is evident in music where pitch classes are organized in reference to pitch center (Podlipniak, 2016) and rhythm measures in reference to musical pulse (London, 2012). By taking into account that both of these features are not present in the vocal expressions of chimpanzees it is reasonable to assume that they were also absent in the vocal repertoire of the common ancestor of chimpanzees and humans and thus are relatively evolutionarily young innovations.

The coexistence of different evolutionarily aged features in our musical expression encourages us to think of human musicality in an analogous way, as in the faculty of language (Hauser et al., 2002) namely in terms of musicality in the broad and narrow sense. While musicality in the broad sense can encompass a set of abilities which develop at an early stage of human ontogenesis, and which allow the identification of pitch contour, changes in tempo, and dynamics as parts of the affective prosody (Zimmermann et al., 2013), only abilities which constitute musicality in the narrow sense enable the recognition of musical structure. From this perspective the actual question about the origin of music is related to the origin of these abilities which constitute musicality in the narrow sense. What makes music a phenomenon distinguishable from other experiences of sounds is the interpretation of sounds by the human nervous system not just as unrelated sound events but as a psychologically unified musical structures. Although it has been proposed that musical structure (e.g., melody) was discovered by early humans due to the resemblance of musical sounds to the acoustic characteristics of human vocalizations (Purves, 2017), the real problem is why the nervous system of early humans started to interpret certain acoustic parameters as discrete pitch and rhythm units organized into syntactically based sequences. After all, despite the fact that a sound stimulus is usually a very complex phenomenon composed of an enormous number of spectral and temporal cues, different species recognize their species-specific vocalizations using only the subsets of features which are distinctive solely to their song cognition (Bregman et al., 2016; Shannon, 2016). In other words, every species is sensitive to specific acoustic cues due to the proclivities of its own nervous system. In addition, speech and music, two complex hierarchical cognitive systems of *H. sapiens* (Fitch, 2014), operate with a restricted number of different distinctive features (Patel, 2008). Therefore, there must be something which predisposes humans to focus their attention on particular acoustic features whilst ignoring others. What is more, the vast majority of musical structures, especially songs of tribal communities (Blacking, 1973), are organized according to certain syntactic rules (Koelsch, 2013 but see London, 2011), which suggests that musical syntax is a natural trait of human vocalization. Apart from this, musical syntax based on pitch classes and rhythmic units measured in reference to musical pulse is a music-specific feature (Fitch, 2013a). This raises the question about the evolution of the abilities which allow the recognition of musical syntax, which seems to be the core of human musicality.

## MUSICAL SYNTAX AS A MUSIC SPECIFIC FEATURE

The human ability to organize and interpret stimuli as syntactically complex sequences is often regarded as a milestone in the evolution of human cognition (Hauser et al., 2002; Fitch, 2014). Even though syntax, understood in the broad sense as rules combining discrete elements into sequences (Patel, 2008), can be attributed to certain animal songs (Okanoya, 2013; Suzuki et al., 2016), both musical and language syntaxes seem to be exceptional in terms of their complexity and function (Fitch and Jarvis, 2013). Musical syntax however, is often thought of as a derivative of language syntax (Patel, 2008 but see Jackendoff and Lerdahl, 2006), or as a product of domain-general structural computation (Fitch, 2014; Van de Cavey and Hartsuiker, 2016) rather than a functionally separate phenomenon. In fact, there are good reasons to assume the general role of structural computation in the processing of language and musical syntaxes. There are many neuroimaging studies that show an overlap in the activation of cortical structures (located in the inferior frontal gyrus) during the processing of musical and language syntaxes (Patel et al., 1998; Maess et al., 2001). Moreover, there is also research which reveals that the same structures (e.g., Broca's area)

play a certain role in the performing of other tasks involving sequential ordering (Tettamanti and Weniger, 2006; Friedrich and Friederici, 2009; Higuchi et al., 2009; Wakita, 2014). Also, the cognitive deficits observed after lesions in the Broca's area include not only deficits in production and recognition of language syntax but also action execution and observation as well as musical syntax processing (Fadiga et al., 2009).

However, from a behavioral point of view, language and music are functionally different phenomena. The former communicates referential meaning by the means of intersubjectively understandable concepts (Bickerton, 2010), whereas the latter exchanges information which is at least more ambiguous in terms of its semantic content (Cross, 2005). Since natural selection operates on phenotypes, we can assume that syntactical language evolved because it gave an advantage to our ancestors over those individuals who could not grasp the syntactic rules present in the language of our ancestors. This means that in the case of cognitive abilities natural selection acts directly upon the behavioral effects of the activity of brain structures. Therefore, when searching for the evolutionary origin of a particular cognitive trait it seems more reasonable to look at it from Tinbergen's perspective (Tinbergen, 1963), namely by considering the possible ultimate function of the brain products (e.g., syntactical language), rather than the proximal function which a particular brain structure fulfills in the processing of a certain task or tasks. From this point of view, the circuitries which perform whole task such as the processing of syntactical language should be assumed as the result of natural selection rather than isolated brain structures which take part in the operations of these circuitries (e.g., Broca's area). Since evolution usually optimizes organisms by the means of adjusting the existing traits to new functions (Jacob, 1977), it is not surprising that the same brain structures are often parts of functionally different circuits. Of course, the human brain is characterized by having huge plasticity, thus the general ability to attribute "tree structures" (Fitch, 2014) (complex syntax) to stimuli, is not restricted solely to language and music. It is well known that people are able to implicitly learn the artificial syntaxes of different stimuli (Reber et al., 1999). Nonetheless, in comparison to artificial grammars, both language and music seem to be exceptional in respect to the rate and easiness of the implicit learning of their syntactic rules by children (Jablonka and Lamb, 2005; Tillmann, 2005). The fact that there is no culture without language and music additionally strengthens the point of view that musicality, similar to language abilities, is a natural part of human behavior rather than being a very old cultural invention similar to writing or playing chess.

## Musical Syntax vs. Language Syntax

There are many similarities between language and musical syntaxes. Both are compositional and hierarchical (Merker, 2002) and both generate long-distance dependencies (Bickerton, 2009; Woolhouse et al., 2016). The default mode of language—speech—is like music in the auditory domain, and it has been observed that the processing of music and speech syntactic tasks activates the peri-Sylvian network which connects the inferior frontal gyrus with sensory cortices located in the temporal lobes (Fitch, 2014). However, musical and language syntaxes are also quite different in many respects. First of all, music is composed of different units to those of language. The basic units of speech are phonemes which are experienced in our internal world as unique qualities hardly comparable with our experience of pitch class. Their discrimination is also based on different spectral cues. Pitch classes are recognized by the fundamental frequency of harmonic sound ($F_0$) (Stainsby and Cross, 2008) whereas phonemes mainly by the spectral and temporal shape of sound (Xu et al., 2005). Although the processing of certain characteristics of spectral shape is important for the discrimination of timbre in music (McAdams and Giordano, 2008), the role of timbre in musical syntax is at least doubtful. Admittedly, timbre can play an important role in the structural organization of music as it is observed in certain musical styles such as in the deep throat singing of Tuva and Mongolia (Levin and Süzükei, 2006), the music of the Jew's harp (Fox, 1988), and tabla music in India (Patel, 2008). There are also musical cultures (e.g., Yakut culture in Siberia) in which the structure-forming function of pitch is extremely reduced whereas timbre seems to be a dominant factor which structures the sound order. Nevertheless, in all these cases timbral structure is hardly comparable to pitch and rhythm structure mainly because of the multidimensional perceptive character of timbre (Lerdahl, 1987). Also, our mental images of timbre in music and phonemes in speech differ, although both are based on the interpretation of the spectral shape of sound. Therefore, even though language grammar (Lerdahl and Jackendoff, 1983), prosodic structure (Heffner and Slevc, 2015) as well as phonotactics and morphonotactics are to some extent comparable to musical syntax (Lerdahl, 2013), the perceptual salience of these phenomena is disparate. The crucial difference between language and musical syntax however is related to the function which these syntaxes fulfill in music and language being different communicative phenomena. In language, syntax is mapped into conceptual meaning (propositional semantics; Hilliard and White, 2009) which allows a concatenation of meaning i.e., putting together two or more units and thereby creating a new meaning in comparison to the meanings of those units alone (Bickerton, 2009). But this process of mapping does not seem to be unidirectional as semantics can influence syntactic rules as in the case of some verbs in which the meaning determines grammatical patterns (Dor and Jablonka, 2000). Thus, the function of language syntax is strictly related to communication of specific conceptual meanings (Dor and Jablonka, 2001). In contrast, the function of syntax in music has nothing to do with such a complex interdependence between syntax and concepts observed in language. Although there is an endless dispute over the existence and character of musical semantics (Patel, 2008; Koelsch, 2013; Reybrouck, 2013; Seifert et al., 2013) even if one admits that music can communicate referential meaning, both the type of this meaning (Dor and Jablonka, 2001) and its relation to musical syntax (Lerdahl, 2013) are definitely different than in language.

## Musical Rhythm and Pitch as the Basis for Musical Syntax

There is currently no agreement about what musical syntax actually is (Patel, 2003; London, 2011; Koelsch, 2013; Lerdahl, 2013; Asano and Boeckx, 2015; Heffner and Slevc, 2015). The

majority of research on musical syntax has been conducted on Western artistic music—especially on the functional relations between chords (Koelsch, 2013). But the case of Western artistic music seems to be an inadequate example of human musical expressions as the manifestation of *H. sapiens* musicality (Jackendoff and Lerdahl, 2006). After all, functional harmony is an exception within the wide variety of world music. Although music based on functional harmony has become more and more widespread in the last century, its history is very young in comparison to the ancient history of music without functional harmony. Nevertheless, the experience of even simple melody without any accompaniment necessitates the recognition of syntactic relations. These relations are hierarchical, meaning that the sequence of sounds is interpreted by the nervous system as being composed of units (sounds perceived as belonging to a particular pitch class and having a particular rhythmic measure) which possess different prominence. This prominence attributed to sounds as elements of metrical (London, 2012; Fitch, 2013b) and pitch (Lerdahl and Jackendoff, 1983; Huron, 2006) patterns is a mental construct as with the prominence of grammatical categories in language. However, in the case of music the metrical or tonal prominence is rather "felt" than "conceptually known" as in language cognition. This preconceptual character of musical hierarchy is especially evident when music is experienced by non-musicians who, in contrast to professional musicians (Burns and Ward, 1978), do not always recognize pitch intervals by the means of categorical perception (Smith et al., 1994). However, even musically trained listeners experience musical hierarchy in such a preconceptual way despite the fact that they are additionally able to recognize musical structure in terms of precise mental categories. What seems to be a source of the preconceptual experience of musical hierarchy is somehow related to motor and emotional brain processes. The recognition of meter in music can be understood as a kind of entrainment which exists in the connection between our auditory and sensorimotor systems (London, 2012). It has even been proposed that human metrical interpretation of music is based on hidden sensorimotor activity (Repp, 2007). This sensorimotor activity during listening to music leads in turns to emotional reactions (Sievers et al., 2013). Also, the recognition of pitch hierarchy causes measurable emotional reactions (Steinbeis et al., 2006; Koelsch et al., 2008; Mikutta et al., 2015), and perception of pitch changes (often described as "leaps" and "steps") can lead to sensorimotor interpretation (Nikolsky, 2015). Since emotion is an evolutionarily old motivational mechanism (Toates, 1988), the function of which is to assess a potential danger or attractiveness of perceived stimuli (Panksepp, 1998), the tight connection between musical structure and emotions suggests the biological importance of this human specific interpretation of sound in terms of pitch and metric hierarchy.

## Musical Syntax and Emotions

During the processing of musical syntax, people experience a set of subtle emotional reactions (emotional qualia) dependent on the position of a note in each syntactic context (Huron, 2006; Margulis, 2014). This observation is supported by the fact that the bilateral Amygdalae and the orbitofrontal cortex are differently activated during listening to music depending on the syntactic relations of musical sequences (Mikutta et al., 2015). In spite of the coexistence of an affective experience during listening to musical syntax, the emotional reaction to musical syntax is often suggested as the mere result of cognitive recognition of syntactic structure (Koelsch et al., 2008) rather than an integral part of this recognition. For example, Huron (Huron, 2006) has proposed that the association of emotions with particular pitch classes is the result of the so called "misattribution effect." In the case of music it is the misattribution of limbic reward or punishment (caused by fulfilling or not fulfilling predictions about which pitch class will be next) to the pitch classes themselves depending on the general mechanism of prediction. However, in the original misattribution effect (Dutton and Aron, 1974) the feeling experienced in response to a stimulus (e.g., the instability of a bridge perceived by the vestibular and visual systems) is misattributed to another stimulus (e.g., a woman perceived by sight). But in Huron's example there is only one stimulus—sound. Because prediction is the ultimate function of the nervous system (Llinás, 2001) one can assume that every perception is based on prediction. The limbic reward (or punishment) in response to well predicted (or falsely predicted) stimuli is an evolutionarily old mechanism of the assessment of stimuli (Panksepp, 1998), inseparable from every perception. In other words, both the emotional reaction to a particular stimulus and the prediction of stimuli are parts of cognition. Therefore, the prediction processes of the nervous system cannot be treated similarly to external stimuli as a source of emotions. The actual source of emotion is an external stimulus and a prediction process is one of the mechanisms, the function of which is to deliver information about the external world that is assessed by emotions. Moreover, if the emotional effect was solely the result of prediction then the emotional reactions to equally predicted stimuli should be the same, independent of whether they are parts of e.g., musical syntax, speech phonotactics or the sequence of timbres in music (Gorzelańczyk et al., 2017). Yet the emotional experience of musical syntactic relations seems qualitatively different from the experience of phonotactics and other syntactically organized sequences. This difference is evident if we compare singing with speech. Although both of these vocal expressions are composed of syntactically organized sounds, the variations of pitch in time occur much slower in singing than in speech (Zatorre and Baum, 2012) and the emotional impact of singing on listeners seems to be greater in comparison with speech from our childhood (Nakata and Trehub, 2004) and lasting throughout our lives.

Taking into account the behavioral specificity of singing, both prediction and emotional reactions to successive sounds are in this case rather the integral parts of a mental tool dedicated to processing musical structure. From this perspective, different subtle emotional reactions to a variety of possible syntactic relations are the elements of a functionally specific form of communication, similar to how semantics is strictly connected with grammar in natural language. In other words, music can be understood as a mapping system in which syntactical relations are mapped into preconceptual emotional cues. For the majority of the human population, the recognition of musical syntax is a solely preconceptual experience. Of course, the

conceptual level of musical syntax recognition is achievable by professional musicians. However, in contrast to implicitly learned tacit syntactic knowledge by non-musicians, this additional level of competence necessitates strenuous explicit learning. Changes in the neural architecture of musicians in response to environmental influences (explicit learning) could perhaps represent a kind of phenotypic adaptation within the cognitive domain which is possible due to the brain's plasticity (Moreno and Bidelman, 2014; Strait and Kraus, 2014). In fact, it has been observed that music performance affects the transcription of genes that are related to dopaminergic neurotransmission, motor behavior, neuronal plasticity, and neurocognitive functions such as learning and memory (Kanduri et al., 2015) which can be responsible for the observed differences between musicians and non-musicians. Therefore, the fact that musical syntax can be recognized by musically trained individuals at a conceptual level shows how much cultural influence can extend cognitive skills rather than saying something about its primordial nature. Since for average humans the syntactic relations in music, both metrical prominence (especially evident when they dance or tap), and pitch prominence (when they sing), are somehow felt but are difficult to express conceptually, it is reasonable to assume that motor, emotional and cognitive processing are an integral part of the ability to recognize and produce syntactically organized music.

## MUSIC AND THE BALDWIN EFFECT

The main problem concerning the evolution of the abilities to use syntax as a part of any communicative system is related to the question about the cause of "syntax genes" proliferation (Dor and Jablonka, 2001). If musical or language syntax is adaptive due to extending communicative capabilities, then this means that it must be used by at least two individuals. After all, as long as one individual cannot use syntax in communication with another individual, the use of syntax by the latter is useless. Even with the use of music in the form of self-communication, as in the case of "personal song" in the musical cultures of Siberia, Far East, and Amerindian tribes (Nikolsky, 2015), the development of the abilities to organize music syntactically necessitates learning from another individual. However, the appearance of a new genetic trait in the population is usually a result of accidental mutation or recombination. Because the probability of the coincidence of identical mutations of the same allele in the same generation is very low the appearance of a genetically based predisposition to organize vocal expression syntactically seems puzzling. In other words, all advantages of syntax are useless in a population in which only one individual is able to produce and recognize syntactically organized sequences. A possible solution to this problem can be the Baldwin effect (Baldwin, 1896a; Simpson, 1953).

### The Baldwin Effect

The Baldwin effect is an evolutionary mechanism which transforms a culturally invented and acquired trait into an instinctive trait by the means of natural selection (Baldwin, 1896a; Simpson, 1953; Hall, 2001). Although this mechanism was independently proposed at the end of the Nineteenth Century by at least three different people (Baldwin, 1896a,b; Morgan, 1896; Osborn, 1896) it was forgotten with some exceptions (Simpson, 1953; Waddington, 1953a,b), for more than half of the next century. Only in the last few decades of the twentieth century has the Baldwinian idea started to inspire scientists and philosophers and has regained popularity (Godfrey-Smith, 2003). The core of this concept is very simple. Some animals due to their cognitive flexibility learn new adaptive behaviors in response to environmental changes. If a particular behavior is adaptive, lasts many generations, and its learning is strenuous and time-consuming, sooner or later a genetically based predisposition appears and starts to be favored by natural selection (Dor and Jablonka, 2000, 2010; Godfrey-Smith, 2003; Jablonka and Lamb, 2005). Therefore, the Baldwin effect is a combination of learning and the genetic assimilation of a learned trait (Dor and Jablonka, 2000, 2001; Godfrey-Smith, 2003). The process of Baldwinian evolution occurs in three stages: (i) the appearance of a new environmental challenge, (ii) the invention of a new behavior as a response to the new environmental challenge and its proliferation by the means of learning—at this stage natural selection favors cognitive plasticity, (iii) the appearance of a new genetically based predisposition (canalization, Jablonka and Lamb, 1995; Dor and Jablonka, 2010)—at this stage natural selection favors less flexible individuals but faster at exhibiting a particular adaptive behavior (Godfrey-Smith, 2003).

Since an important factor in the Baldwinian evolution of human behavior is the social character of our species, the aforesaid environmental challenge can be a part of the socio-cultural products of hominins. This specific socio-cultural environment often described as a "cultural niche" (Godfrey-Smith, 2003) has been proposed as a crucial element in the evolution of natural language (Bickerton, 2010; Deacon, 2010). In some regards, the evolution of language can represent a niche construction (Deacon, 1997, 2003)—a kind of niche extension or elaboration. The importance of this "cultural niche" seems evident when taking into account that at least during the last 2 million years hominins have become more and more socially complex animals in comparison to other primates (Dunbar, 2014). Living in a complex social group definitely causes new challenges which influence both natural selection and niche construction. In the process of niche construction hominin brains and hominin culture can be considered as a specific environment in which language evolved (Deacon, 2003). From this perspective the Baldwinian process is a part of the gene-culture coevolution (Lumsden and Wilson, 1982; Richerson et al., 2010; Gintis, 2011). Since speech (the "default mode" of language) is similar to music in many respects (both are transmitted by the acoustic domain, both are syntactical systems composed of discrete elements etc.), and language is assumed to be a crucial factor in the cultural evolution of *H. sapiens*, it seems reasonable to assume that the evolution of human musicality is somehow related to the evolution of language. However, language is a very elaborate signal related to the exchange of conceptual meaning. The presence of non-symbolic and non-conceptual culture in many other species (Cantor and Whitehead, 2013; van de Waal et al., 2013; Fehér et al., 2016) indicates that the beginning of

cultural niche construction can be based on the exchange of preconceptual meaning. Therefore, music as an example of a communication system operating on preconceptual meaning is a good candidate to be a part of more ancient communicative tool other than language and so the proposed Baldwinian scenarios of language origin must differ from the possible Baldwinian processes that led to the emergence of music.

## The Baldwinian Evolution of Music

Human musicality seems to be a very good example of the potential effects of Baldwinian evolution (Podlipniak, 2015). Huron has suggested that apart from certain reflexes, the human auditory experience is mainly influenced by learning (Huron, 2006). The predominance of learning in shaping human auditory cognition implies, according to Huron, that the auditory environment of hominins must have been very semiotically unstable. Following Huron's reasoning, such a semiotic instability led to the great variety of music found around the world. However, music as a product of human musicality is characterized not only by culture-specific features but also by universals (Nettl, 2000; Bispham, 2009; Brown and Jordania, 2011; Savage et al., 2015) which suggests that apart from the cultural (environmental) influence, the music-specific genetic constraints also shape the musical mind of every human. This means that on the one hand, musicality develops spontaneously and effortlessly but on the other hand, the learning of more sophisticated musical skills, as in the case of professional musicians, is time-consuming and necessitates a lot of effort. The transmission of musical information has its roots in the human ability of vocal learning which is exceptional among primates (Janik and Slater, 1997; Fitch and Jarvis, 2013). However, human vocal learning is canalized into an imitation of selected acoustic characteristics rather than the literal copying of every heard sound (Jackendoff and Lerdahl, 2006). People are very skillful at imitating the distinctive features of phonemes, the temporal order of sound sequences, and fundamental frequency of harmonic sounds but not very talented when they try to simulate the barking of a dog or environmental sounds such as the noise of a refrigerator which seem a very simple task for many parrots. This canalization suggests that apart from the aforementioned environmental instability, certain circumstances related to hominin vocal expressions must have been stabilized long enough during numerous generations to cause natural selection to have promoted an instinct to learn only selected sound features (Briscoe, 2000; Gibson and Tallerman, 2011). As a result, speech and music, similar to many songbirds' songs, are examples of so called ritual culture (Merker, 2009). The most important characteristic of ritual culture is its transmission by the means of imitative social learning (Merker, 2005, 2012). In contrast to non-imitative social learning, learning by imitation consists of copying the behavior of other individuals (Jablonka and Lamb, 2005). Therefore, what is important in the transmission of ritual is not the result of a particular action but the action itself (Merker, 2009). In the case of music, transmitted units are pitch classes and rhythm measures (Merker, 2002, 2003). After all, a melody is recognized independent of whether it is played slower or faster on the flute, piano, or when sung. What is important for the recognition of melodic pattern is its pitch and rhythm structure, not timbre or dynamics. In this respect music seems to be an even more striking example of ritual culture than speech in which, apart from poetry, what is crucial is the transmission of the semantic content of utterance and not its literal form. However, the thing that makes melody easier to remember is musical syntax, the evolution of which is most easily explained by the Baldwin effect.

The Baldwin effect may promote a particular trait due to different adaptive functions. For example, Morgan proposed that organic evolution (the term which Morgan used to describe the process known today as the Baldwin effect) can explain the origin of bird songs which evolved as a result of sexual selection (Morgan, 1891, 1920). Since bird songs are similar to human music in many respects, it is tempting to explain the origin of musical syntax by the means of the Baldwin effect in which the adaptive function of music is to attract sexual partners. If this is true, musicality could have been used by hominins as a mating handicap (a mark of the quality of a mate, Zahavi, 1975; Zahavi and Zahavi, 1997; Miller G. F., 2000) since the production and recognition of musical syntax is costly in terms of energy (necessary to process the perceived sounds and to control the vocal production of songs) and time spent on singing (which for example can be used for foraging instead). In such a scenario the syntactical complexity of a hominin song should attract females more than a song that lacks such a complexity due to the costliness of the song's complex structure being an indicator of fitness (Miller G., 2000). If these female preferences had been stable enough throughout many generations, the Baldwinian mechanism should have transformed the learning of culturally invented rules of musical structure into an instinct to learn and organize musical sounds in a syntactic way. This emerged instinct to learn the distribution of music-specific discrete elements based on intuitive recognition of their probability of occurrence would have left space for idiosyncratic song modifications similar to those observed in songbirds' behavior. Such a leeway in creating new songs allows the sustaining of the process of sexual selection based on the songs' complexity. However, a study of female preferences toward musical complexity showed that women do not have a tendency to prefer more complex music during and around ovulation (Charlton et al., 2012), which does not support the Baldwinian scenario of music origin based on sexual selection. Similarly, research shows that musical aptitude and achievements are not a predictor of mating success (Mosing et al., 2014). Of course these studies are not conclusive and more studies are necessary to test a possible role of sexual selection in music evolution. Nevertheless, so far the Baldwinian scenario in the sexual selection of human musicality needs more empirical support to be convincing.

Another possible scenario of the Baldwinian origin of music is related to the idea that music can serve as a tool of social consolidation. This scenario started the moment a new social challenge first appeared. The increasing size of the hominin population caused an increase in inter-individual and inter-group competition for food and other resources (Dunbar, 2014). One way to cope with this problem was to form alliances between individuals belonging to a group. This strategy has been

observed in other primates, including our closest relatives—the chimpanzee (Mitani, 2009; Gilby et al., 2013), which suggests that hominins could use a similar strategy. Dunbar has proposed that as group size increased, grooming as the main tool to sustain social alliances became insufficient. Instead, hominin vocalizations started to serve as a tool of social consolidation (Dunbar, 1996). While this idea seems to be unconvincing as far as the origin of language is concerned (Dunbar and Lehmann, 2013; Grueter et al., 2013) its validity as an explanation of the origin of music still remains an open question. An increasing number of studies suggest that communal singing can facilitate social bonds (Dunbar et al., 2012; Tarr et al., 2014; Pearce et al., 2015, 2016, 2017), which supports Dunbars' idea. However, the precise mechanism of how music could act in this way remains a puzzle. Although the obtained results of the aforementioned studies must not necessarily be the effects of the adaptive value of social bonding, being for example a byproduct of sexually selected behavior, certain characteristics of musical syntax seem to bespeak the hypothesis of music as a tool of social consolidation. A big problem which afflicts individuals living in groups is the situation in which some individuals use resources obtained by other individuals—the so called "free-riding problem." Communal song rituals demand that all participants must know the musical structure of the song. The learning of a particular song's structure is time-consuming and necessitates strenuous imitation of the vocal behavior of others especially in the case when hominins did not possess an instinct to learn musical syntax. Therefore, by devoting equal effort in order to learn a ritualized song, communal singing can serve as a good test of being prone to act together with others. After all, poor singers can be easily recognized which can lead to ostracism. In other words, the consolidation effect observed after communal singing can be a product of the unconscious assessment of other individuals in terms of their proclivity to being free-riders. From this perspective, a lack of synchronization hinders consolidation which can be a result of detecting potential free-riders and can lead to looking for new allies.

Also a proximal explanation of the mechanism responsible for the consolidating power of music can be related to observed characteristics of music processing by the human nervous system. It is possible that music consolidates individuals by the means of temporal and spectral synchronization between the brain states of co-performers (Bharucha et al., 2011). If this is true, our predecessors had to simultaneously imitate their vocalizations in order to sustain social trust (Podlipniak, 2016). This collective imitation became the beginning of a consolidating vocal ritual. Without any predisposition which canalized vocal learning so that hominins would have been sensitive to certain acoustic features, the process of the learning of vocal rituals would have been very strenuous and time-consuming. During this time, the second stage of Baldwinian evolution began in which natural selection preferred individuals who were characterized by the most flexible learning. In order to learn new melodies and sing them together then the appropriate predictions of what (which particular pitch class) and when (the position of a particular rhythm measure in relation to musical pulse) would happen in the near future was necessary. Syntax is exactly what makes the successful predictions of sound events during singing easier and, as a consequence, facilitates collective singing. At this stage, the learning of simple syntactic rules would have been accessible to hominins in a similar way to people that learn artificial syntaxes today. Because the costs of ritual learning were high an individual who was accidentally endowed with proclivities to predict the melody better than others gained an advantage over the rest of a group. In the long run, the progeny of this individual has dominated the whole population.

In a similar vein the Baldwin effect could have contributed to the origin of music if its adaptive function advertised the defending skills of a group (Hagen and Bryant, 2003; Hagen and Hammerstein, 2009; Jordania, 2014). However, in case an acoustic aposematism (instrumental music, singing) had been directed against predators (Jordania, 2014) a possible role of the Baldwinian mechanism in the origins of human musicality would have been restricted solely to the canalization of the elements of musical display which are recognizable by predators. These elements are a part of musicality in a broad sense such as pitch contour, changes in tempo, and dynamics rather than the syntactic relations specific to the discrete structure of human music. In the Baldwinian scenario, the initially invented complexity of musical syntax which became a part of the hominin cultural niche was most probably accessible (comprehensible) only to the hominin species. Therefore, predator reactions did not depend on the subtleties of musical structure and had not been a selective factor which could have influenced the process of canalization of musical-syntactical abilities which actually define musicality in a narrow sense. In contrast, if music was a coalition signaling display directed toward conspecifics (Hagen and Bryant, 2003), the Baldwinian scenario could have been very similar to that presented above in the case of consolidation as the adaptive function of music. The only difference is that this time the selective pressure which would have been responsible for the canalization of human musicality had been induced by the reactions of enemies. However, while rhythm syntax seems to contribute to the signaling/deteriorating function of music (despite the fact that people from different cultures can recognize different units of musical pulse in the same piece of music (London, 2012), a well synchronized rhythm is perceivable independent of whether the rules of rhythm syntax are familiar to us or not), the origin of pitch syntax is more problematic as a result of its possible adaptive signaling/deteriorating function.

First of all, even today when people are most probably endowed with an instinct to learn pitch system and pitch syntax, well spectrally synchronized music which is based on an unfamiliar pitch system can be perceived as being out-of-tune (Ellis, 1885), which can be a sign of a poor performance rather than a signal of a performers' coalition or consolidation. Additionally, the recognition of pitch syntax by contemporary humans depends on tacit knowledge about the statistical distribution of pitch classes in a particular music (Tillmann et al., 2000; Tillmann, 2005; Huron, 2006). Foreigners who listen to unfamiliar music usually experience different tonal qualia than people familiar with this music (Castellano et al., 1984; Kessler et al., 1984; Stevens, 2004, 2012; Curtis and Bharucha, 2009). Although this situation looks similar to the

aforesaid difference in musical pulse perception, it differs in one important aspect. Pitch syntax is based on pitch hierarchy in which the most prominent place is pitch center, the experience of which is accompanied by the emotional qualia of completeness, resolution etc. There is nothing resembling pitch center in rhythm hierarchy. The misrecognition of actual tonal relations in reference to pitch center by foreigners can lead to divergence between the observed emotional expression of performers (also by the means of expressive dynamics) and tonal qualia felt by those foreigners. Such a divergence can also be a signal of poor performance. Importantly, without the canalized strategy to learn pitch syntax the differences in musical dialects between hominin groups would have been even greater than between modern geographically distant musical cultures causing the aforesaid divergence to be even greater. Therefore, while in the scenario of music origin in which music is a coalition signaling system the Baldwinian mechanism can explain the origin of musical rhythm so it is difficult to imagine a similar role of the Baldwin effect in the origin of pitch syntax as a result of its coalition signaling function.

It is worth mentioning however, that the possible different adaptive functions of music are not mutually exclusive and the Baldwin effect could have played an important role at different stages of the gradual process of the evolution of human musicality. Nevertheless, it seems that Baldwinian evolution was necessary at least in the process that led to the emergence of the complex musical syntax as a part of hominins' singing behavior.

## THE EVOLUTION OF NEW CIRCUITRY

The appearance of ritualized singing behavior among hominins required the development of new abilities and recruitment of existing skills. An important ability which had to be a necessary condition of the development of culturally variable vocal communication is the aforesaid vocal learning (Janik and Slater, 1997). This ability had to be present at the first stage of the Baldwinian evolution of human musicality. The similarity of vocal pathways in vocal learning birds to cortical–basal ganglia–thalamic–cortical loops in humans suggests the role of the latter in the processing of speech (Jarvis, 2007). In fact, there is an increasing number of studies which emphasize the role of the basal ganglia in the processing of language (Booth et al., 2005) especially in the learning of language during childhood (Krishnan et al., 2016). The fact that language impairment can be the result of neurodevelopmental deficits of the corticostriatal loops (Krishnan et al., 2016) shows that corticostriatal connectivity might have been an important element of evolutionary change leading to the evolution of vocal communication among our predecessors. It is not surprising that the basal ganglia contributes to the processing of sound sequences specific to vocal communication. It is known that the basal ganglia is important in reinforcement learning (Bar-Gad et al., 2003) which is necessary to acquire the majority of culturally transmitted information. This ability is strictly connected to predictive functions which are related to internal timing (Dreher and Grafman, 2002). It is reasonable to assume that in the beginning hominin vocal communication was composed of simple vocal expressions which must have been vocally learned.

In the process of vocal learning the recognition and prediction of distinctive acoustic features is necessary and so the specific connections between the basal ganglia and auditory cortices had to be favored by natural selection during the second stage of the Baldwinian evolution of human musicality. Because contemporary humans are characterized by the ability to sustain and reproduce $F_0$ which is crucial for singing (Bannan, 2012) but not for speaking, the sequencing of sounds based on their $F_0$ characteristics must have become an important trait of hominin sound rituals. The consolidating function of them did not necessitate any referential conceptual meaning. Instead simple emotional cues were enough to establish close social relations. The emotional reinforcement of a well predicted sound i.e., the sound which is perceived as possessing a particular pitch and which happens at an exactly predicted point of time, started to act as a cue for social acceptance. The positive emotional reaction in response to music is in fact an evolutionarily old clue of social acceptance. It has been observed that this emotional reward occurring in response to music is related to the corticostriatal interactions involving auditory cortices and the nucleus accumbens (Salimpoor et al., 2013). Also the perception of beat in music is based on corticostriatal interactions (Grahn and Rowe, 2009, 2013). It is suspected that the role of the cortico-basal ganglia circuits in speech evolution is related to the positive selection of the FOXP2 gene variant (Enard, 2011). However, because the mutation of FOXP2 also impairs rhythm processing in music leaving intact pitch processing (Alcock et al., 2000; Tan et al., 2014) the evolution of the abilities to recognize pitch syntax must have been related to other genetic factors. Independent of what particular genetic factors influence pitch and rhythm processing in music (Tan et al., 2014), human perceptive preferences to recognize pitch classes and rhythm measures as parts of syntactically organized sequences suggest that at the last stage of the Baldwinian scenarios natural selection started to prefer individuals endowed with these canalized perceptive proclivities.

## CONCLUSION

The proposed possible Baldwinian scenarios of the evolution of human musicality solve the problem of the specificity of musical syntax which is functionally and structurally different from language syntax. This specificity suggests that all theories which explain the syntactic characteristic of music as a byproduct seem unconvincing. After all, music seems to be the only spontaneously emerging syntactic system apart from speech, dance (Opacic et al., 2009), and some drum (Winter, 2014) and whistled languages (Carreiras et al., 2005; Güntürkün et al., 2015; Meyer and Busnel, 2015). Although the details of these proposed scenarios are speculative, future research can elucidate which particular elements of these scenarios are more probable. The most promising would be studies which concentrate on the comparison between the functions of cortico-striatal loops (Gorzelańczyk, 2011) in speech and music processing. In order

to investigate which particular loop is mostly involved in the processing of musical syntax, neuroimaging studies could be conducted in which the activity of limbic and dorsolateral-prefrontal loops can be compared during performing tasks related to the recognition of musical and language syntaxes. Although the complex interactions between genetic, epigenetic, and cultural information which occur in evolution have not so far been explained in detail (Jablonka and Lamb, 2005) it seems important to consider them in future models of the evolution of human musicality. The rapidly advancing development of genomics and transcriptomics allows us to expect that the details of these complex interactions will be better understood in the near future.

Another possible study which could be conducted to test the proposed Baldwinian origin of musicality is to compare the results of singing syntactically simple with syntactically complex (more demanding in terms of explicit learning) tonal melodies. If singing syntactically complex melodies leads to a greater consolidation of singers or if the singing group is assessed by others as being more consolidated than singers of the syntactically simple melodies then it would suggest that the tendency which was proposed as the main source of the Baldwinian evolution of musicality is still present in the human population. Also the comparison of singing tonal melodies by a group of people with other collective sound expressions such as simultaneously reading prose, reciting poetry, drumming, and singing atonal melodies in free rhythm (without musical syntax) would be informative as far as the question of what particular features of human sound expressions are responsible for the observed effects. If the proposed consolidating function of human musicality is tenable, then the consolidating effect of singing tonal melodies and drumming should be greater than other collective behaviors. Nevertheless, in order to better understand the origin of music the broad holistic view of human musicality is necessary. This view should be based on integrated knowledge taken from such disciplines as genetics, evolutionary biology, paleoanthropology, neuroscience, psychology, archeology, ethnomusicology and cognitive musicology.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## REFERENCES


Alcock, K. J., Passingham, R. E., Watkins, K., and Vargha-Khadem, F. (2000). Pitch and timing abilities in inherited speech and language impairment. *Brain Lang.* 75, 34–46. doi: 10.1006/brln.2000.2323

Asano, R., and Boeckx, C. (2015). Syntax in language and music: what is the right level of comparison? *Front. Psychol.* 6:942. doi: 10.3389/fpsyg.2015.00942

Baldwin, J. M. (1896a). A new factor in evolution. *Am. Nat.* 30, 441–451. doi: 10.1086/276408

Baldwin, J. M. (1896b). A new factor in evolution (Continued). *Am. Nat.* 30, 536–553. doi: 10.1086/276428

Bannan, N. (2012). "Harmony and its role in human evolution," in *Music, Language, and Human Evolution*, ed N. Bannan (Oxford: Oxford University Press), 288–340.

Bar-Gad, I., Morris, G., and Bergman, H. (2003). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Prog. Neurobiol.* 71, 439–473. doi: 10.1016/j.pneurobio.2003.12.001

Bharucha, J., Curtis, M., and Paroo, K. (2011). "Musical communication as alignment of brain states," in *Language and Music as Cognitive Systems,* eds P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Oxford; New York, NY: Oxford University Press), 139–155.

Bickerton, D. (2009). "Syntax for Non-syntacticians a brief primer," in *Biological Foundations and Origin of Syntax*, eds D. Bickerton and E. Szathmáry (Cambridge; London: The MIT Press), 3–13.

Bickerton, D. (2010). *Adam's Tongue: How Humans Made Language, How Language Made Humans*. New York, NY: Hill and Wang.

Bispham, J. C. (2009). Music's "design features": musical motivation, musical pulse, and musical pitch. *Music. Sci.* 13, 41–61. doi: 10.1177/1029864909013002041

Blacking, J. (1973). *How Musical is Man?* Seattle, WC; London: University of Washington Press.

Booth, J. R., Wood, L., Lu, D., Houk, J. C., Bitan, T., Kotz, S. A., et al. (2005). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Brain* 45, 392–399. doi: 10.1016/j.tics.2010.06.005

Bregman, M. R., Patel, A. D., and Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proc. Natl. Acad. Sci. U.S.A.* 113, 1666–1671. doi: 10.1073/pnas.1515380113

Briscoe, T. (2000). Grammatical acquisition: inductive bias and coevolution of language and the language acquisition device. *Language* 76, 245–296. doi: 10.2307/417657

Brown, S., and Jordania, J. (2011). Universals in the world's musics. *Psychol. Music* 41, 229–248. doi: 10.1177/0305735611425896

Burns, E. M., and Ward, W. D. (1978). Categorical perception–phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *J. Acoust. Soc. Am.* 63, 456–468. doi: 10.1121/1.381737

Cantor, M., and Whitehead, H. (2013). The interplay between social networks and culture: theoretically and among whales and dolphins. The interplay between social networks and culture: theoretically and among whales and dolphins. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368, 1–10. doi: 10.1098/rstb.2012.0340

Carreiras, M., Lopez, J., Rivero, F., and Corina, D. (2005). Linguistic perception: neural processing of a whistled language. *Nature* 433, 31–32. doi: 10.1038/433031a

Castellano, M. A., Bharucha, J. J., and Krumhansl, C. L. (1984). Tonal hierarchies in the music of north India. *J. Exp. Psychol. Gen.* 113, 394–412. doi: 10.1037/0096-3445.113.3.394

Charlton, B. D., Filippi, P., Fitch, W. T., Hamilton, C., and Perrett, D. (2012). Do women prefer more complex music around ovulation? *PLoS ONE* 7:e35626. doi: 10.1371/journal.pone.0035626

Cross, I. (2005). "Music and meaning, ambiguity and evolution," in *Musical Communication*, eds D. Miell, R. MacDonald, and D. J. Hargreaves (Oxford; New York, NY: Oxford University Press), 27–44.

Cross, I., and Morley, I. (2008). "The evolution of music: theories, definitions and the nature of the evidence," in *Communicative Musicality*, eds S. Malloch and C. Trevarthen (Oxford; New York, NY: Oxford University Press), 61–82.

Curtis, M. E., and Bharucha, J. J. (2009). Memory and musical expectation for tones in cultural context. *Music Percept.* 26, 365–375. doi: 10.1525/mp.2009.26.4.365

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*, *1st Edn.* London: John Murray.

Deacon, T. W. (1997). *The Symbolic Species: The Evolution of Language and the Brain*. New York, NY; London: W.W. Norton.

Deacon, T. W. (2003). "Multilevel selection in a complex adaptive system: The problem of language origins," in *Evolution and Learning: The Baldwin Effect Reconsidered. Life and Mind,* eds B. H. Weber and D. J. Depew (Cambridge; London: Bradford Book; MIT Press), 81–106.

Deacon, T. W. (2010). A role for relaxed selection in the evolution of the language capacity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9000–9006. doi: 10.1073/pnas.0914624107

Dissanayake, E. (2008). If music is the food of love, what about survival and reproductive success? *Music. Sci.* 12, 169–195. doi: 10.1177/1029864908012001081

Dor, D., and Jablonka, E. (2000). From cultural selection to genetic selection: a framework for the evolution of language. *Selection* 1, 33–56. doi: 10.1556/Select.1.2000.1-3.5

Dor, D., and Jablonka, E. (2001). "How language changed the genes: toward an explicit account of the evolution of language," in *New Essays on the Origin of Language* eds J. Trabant and S. Ward (Berlin; New York, NY: De Gruyter Mouton), 149–176.

Dor, D., and Jablonka, E. (2010). "Plasticity and canalization in the evolution of linguistic communication: an evolutionary developmental approach," in *The Evolution of Human Language*, eds R. K. Larson, V. Deprez, and H. Yamakido (Cambridge: Cambridge University Press), 135–147.

Dreher, J.-C., and Grafman, J. (2002). The roles of the cerebellum and basal ganglia in timing and error prediction. *Eur. J. Neurosci.* 16, 1609–1619. doi: 10.1046/j.1460-9568.2002.02212.x

Dunbar, R. I. M. (1996). *Grooming, Gossip, and the Evolution of Language*. London: Harvard University Press.

Dunbar, R. I. M. (2014). *Human Evolution*. London: Penguin books Ltd.

Dunbar, R. I. M., and Lehmann, J. (2013). Grooming and social cohesion in primates: a comment on Grueter et al. *Evol. Hum. Behav.* 34, 453–455. doi: 10.1016/j.evolhumbehav.2013.08.003

Dunbar, R., Kaskatis, K., MacDonald, I., and Barra, V. (2012). Performance of music elevates pain threshold and positive affect: implications for the evolutionary function of music. *Evol. Psychol.* 10, 688–702. doi: 10.1177/147470491201000403

Dutton, D. G., and Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *J. Pers. Soc. Psychol.* 30, 510–517. doi: 10.1037/h0037031

Ellis, A. (1885). On the musical scales of various nations. *J. Soc. Arts* 33, 485–527.

Enard, W. (2011). FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. *Curr. Opin. Neurobiol.* 21, 415–424. doi: 10.1016/j.conb.2011.04.008

Fadiga, L., Craighero, L., and D'Ausilio, A. (2009). Broca's area in language, action, and music. *Ann. N.Y. Acad. Sci.* 1169, 448–458. doi: 10.1111/j.1749-6632.2009.04582.x.

Fehér, O., Ljubičić, I., Suzuki, K., Okanoya, K., and Tchernichovski, O. (2016). Statistical learning in songbirds: from self-tutoring to song culture. *Philos. Trans. R. Soc. London B Biol. Sci.* 372:20160053. doi: 10.1098/rstb.2016.0053

Fitch, W. T. (2006). On the biology and evolution of music. *Music Percept.* 24, 85–88. doi: 10.1525/mp.2006.24.1.85

Fitch, W. T. (2013a). "Musical Protolanguage: Darwin's theory of language evolution revisited," in *Birdsong, Speech, and Language : Exploring the Evolution of Mind and Brain*, eds J. J. Bolhuis and M. Everaert (Cambridge; London: The MIT Press), 489–503.

Fitch, W. T. (2013b). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Front. Syst. Neurosci.* 7:68. doi: 10.3389/fnsys.2013.00068

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys. Life Rev.* 11, 329–364. doi: 10.1016/j.plrev.2014.04.005

Fitch, W. T. (2015). Four principles of bio-musicology. *Philos. Trans. R. Soc. London B Biol. Sci.* 370:20140091. doi: 10.1098/rstb.2014.0091

Fitch, W. T., and Jarvis, E. D. (2013). "Birdsong and other animal models for human speech, song, and vocal learning," in *Language, Music and the Brain*, ed M. A. Arbib (Cambridge; London: The MIT Press), 499–539.

Fox, L. (1988). *The Jew's Harp: a Comprehensive Anthology*. London: Associated University Presses.

Friedrich, R., and Friederici, A. D. (2009). Mathematical logic in the human brain: syntax. *PLoS ONE* 4:e5599. doi: 10.1371/journal.pone.0005599

Gibson, K. R., and Tallerman, M. (2011). "Introduction to Part II: the biology of language evolution: anatomy, genetics and neurology," in *The Oxford Handbook of Language Evolution*, eds K. Gibson and M. Tallerman (Oxford; New York, NY: Oxford University Press), 133–142.

Gilby, I. C., Brent, L. J. N., Wroblewski, E. E., Rudicell, R. S., Hahn, B. H., Goodall, J., et al. (2013). Fitness benefits of coalitionary aggression in male chimpanzees. *Behav. Ecol. Sociobiol.* 67, 373–381. doi: 10.1007/s00265-012-1457-6

Gintis, H. (2011). Gene-culture coevolution and the nature of human sociality. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 366, 878–888. doi: 10.1098/rstb.2010.0310

Godfrey-Smith, P. (2003). "Between Baldwin Skepticisim and Baldwin Boosterism," in *Evolution and Learning: The Baldwin Effect Reconsidered*, eds B. H. Weber and D. J. Depew (Cambridge: The MIT Press), 53–67.

Gorzelańczyk, E. J. (2011). "Functional Anatomy, Physiology and Clinical Aspects of Basal Ganglia," in *Neuroimaging for Clinicians - Combining Research and Practice,* ed J. F. P. Peres (Rijeka: InTech), 89–106.

Gorzelańczyk, E. J., Podlipniak, P., Walecki, P., Karpiński, M., and Tarnowska, E. (2017). Pitch Syntax violations are linked to greater skin conductance changes, relative to timbral violations – the predictive role of the reward system in perspective of cortico–subcortical loops. *Front. Psychol.* 8:586. doi: 10.3389/fpsyg.2017.00586

Grahn, J. A., and Rowe, J. B. (2009). Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J. Neurosci.* 29, 7540–7548. doi: 10.1523/JNEUROSCI.2018-08.2009

Grahn, J. A., and Rowe, J. B. (2013). Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. *Cereb. Cortex* 23, 913–921. doi: 10.1093/cercor/bhs083

Grueter, C. C., Bissonnette, A., Isler, K., and van Schaik, C. P. (2013). Grooming and group cohesion in primates: implications for the evolution of language. *Evol. Hum. Behav.* 34, 61–68. doi: 10.1016/j.evolhumbehav.2012.09.004

Güntürkün, O., Güntürkün, M., and Hahn, C. (2015). Whistled Turkish alters language asymmetries. *Curr. Biol.* 25, R706–R708. doi: 10.1016/j.cub.2015.06.067

Hagen, E. H., and Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Hum. Nat.* 14, 21–51. doi: 10.1007/s12110-003-1015-z

Hagen, E. H., and Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. *Music. Sci.* 13, 291–320. doi: 10.1177/1029864909013002131

Hall, B. K. (2001). Organic selection: proximate environmental effects on the evolution of morphology and behaviour. *Biol. Philos.* 16, 215–237. doi: 10.1023/A:1006773408919

Harvey, A. R. (2017). *Music, Evolution, and the Harmony of Souls*. Oxford: Oxford University Press.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569

Heffner, C. C., and Slevc, L. R. (2015). Prosodic structure as a parallel to musical structure. *Front. Psychol.* 6:1962. doi: 10.3389/fpsyg.2015.01962

Higuchi, S., Chaminade, T., Imamizu, H., and Kawato, M. (2009). Shared neural correlates for language and tool use in Broca's area. *Neuroreport* 20, 1376–1381. doi: 10.1097/WNR.0b013e3283315570

Hilliard, A. T., and White, S. A. (2009). "Possible Precursors of Syntactic Components in Other Species," in *Biological Foundations and Origin of Syntax*, eds D. Bickerton and E. Szathmáry (Cambridge; London: The MIT Press), 161–183.

Huron, D. B. (2006). *Sweet Anticipation : Music and the Psychology of Expectation*. Cambridge; London: The MIT Press.

Jablonka, E., and Lamb, M. J. (1995). *Epigenetic Inheritance and Evolution : the Lamarckian Dimension*. Oxford; New York, NY; Tokyo: Oxford University Press.

Jablonka, E., and Lamb, M. J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge; London: MIT Press.

Jackendoff, R., and Lerdahl, F. (2006). The capacity for music: what is it, and what's special about it? *Cognition* 100, 33–72. doi: 10.1016/j.cognition.2005.11.005

Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134

Janik, V. M., and Slater, P. J. B. (1997). Vocal learning in mammals. *Adv. Study Behav.* 26, 59–99. doi: 10.1016/S0065-3454(08)60377-0

Jarvis, E. D. (2007). Neural systems for vocal learning in birds and humans: a synopsis. *J. Ornithol.* 148, 35–44. doi: 10.1007/s10336-007-0243-0

Jordania, J. (2014). *Tigers, Lions and Humans: History of Rivalry, Conflict, Reverence and Love*. Tbilisi: Logos.

Kanduri, C., Kuusi, T., Ahvenainen, M., Philips, A. K., Lähdesmäki, H., and Järvelä, I. (2015). The effect of music performance on the transcriptome of professional musicians. *Sci. Rep.* 5:9506. doi: 10.1038/srep09506

Kessler, E. J., Hansen, C., and Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and in the west. *Music Percept.* 2, 131–165. doi: 10.2307/40285289

Koelsch, S. (2013). *Brain and Music*. Oxford: Wiley-Blackwell.

Koelsch, S., Kilches, S., Steinbeis, N., and Schelinski, S. (2008). Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. *PLoS ONE* 3:e2631. doi: 10.1371/journal.pone.0002631

Krishnan, S., Watkins, K. E., and Bishop, D. V. M. (2016). Neurobiological basis of language learning difficulties. *Trends Cogn. Sci.* 20, 701–714. doi: 10.1016/j.tics.2016.06.012

Lerdahl, F. (1987). Timbral hierarchies. *Contemp. Music Rev.* 2, 135–160. doi: 10.1080/07494468708567056

Lerdahl, F. (2013). "Musical syntax and its relation to linguistic syntax," in *Language, Music, and the Brain*, ed M. A. Arbib (Cambridge; London: The MIT Press), 257–272.

Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge; London: The MIT Press.

Levin, T. C., and Süzükei, V. (2006). *Where Rivers and Mountains Sing : Sound, music, and Nomadism in Tuva and Beyond*. Bloomington: Indiana University Press.

Llinás, R. R. (2001). *I of the Vortex: from Neurons to Self*. Cambridge; London: MIT Press.

London, J. (2011). "Schemas, not syntax: a reply to Patel," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Oxford; New York, NY: Oxford University Press), 242–247.

London, J. (2012). Three things linguists need to know about rhythm and time in music. *Empir. Musicol. Rev.* 7, 5–11. doi: 10.18061/1811/52973

Lumsden, C. J., and Wilson, E. O. (1982). Précis of genes, mind, and culture. *Behav. Brain Sci.* 5, 1–37. doi: 10.1017/S0140525X00010128

Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* 4, 540–545. doi: 10.1038/87502

Margulis, E. H. (2014). *On Repeat : how Music Plays the Mind*. Oxford; New York, NY: Oxford University Press.

McAdams, S., and Giordano, B. L. (2008). "The perception of musical timbre," in *The Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford; New York, NY: Oxford University Press), 72–80.

Merker, B. (2002). Music: the missing humboldt system. *Music. Sci.* 6, 3–21. doi: 10.1177/102986490200600101

Merker, B. (2003). "Is there a biology of music? And why does it matter?" in *Proceedings of the 5th Triennial ESCOM Conference*, eds C. Kopiez, R. Lehmann, A. C. Wolther, and I. Wolf (Hanover: Hanover University of Music and Drama), 402–405.

Merker, B. (2005). The conformal motive in birdsong, music, and language: an introduction. *Ann. N.Y. Acad. Sci.* 1060, 17–28. doi: 10.1196/annals.1360.003

Merker, B. (2009). "Ritual foundations of human uniqueness," in *Communicative Musicality. Exploring the Basis of Human Companionship*, eds C. Malloch and S. Trevarthen (Oxford; New York, NY: Oxford University Press), 45–59.

Merker, B. (2012). "The Vocal Learning Constellation," in *Music, Language, and Human Evolution*, ed N. Bannan (London: O*xford University Press)*, 215–260.

Meyer, J., and Busnel, R. G. (2015). *Whistled Languages : A Worldwide Inquiry on Human Whistled Speech*. New York, NY; London: Springer-Verlag.

Mikutta, C. A., Dürschmid, S., Bean, N., Lehne, M., Lubell, J., Altorfer, A., et al. (2015). Amygdala and orbitofrontal engagement in breach and resolution of expectancy: a case study. *Psychomusicol. Music Mind Brain* 25, 357–365. doi: 10.1037/pmu0000121

Miller, G. (2000). Mental traits as fitness indicators. Expanding evolutionary psychology's adaptationism. *Ann. N.Y. Acad. Sci.* 907, 62–74. doi: 10.1111/j.1749-6632.2000.tb06616.x

Miller, G. F. (2000). "Evolution of human music through sexual selection," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge: The MIT Press), 329–360.

Mitani, J. C. (2009). Cooperation and competition in chimpanzees: current understanding and future challenges. *Evol. Anthropol. Issues News Rev.* 18, 215–227. doi: 10.1002/evan.20229

Moreno, S., and Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear. Res.* 308, 84–97. doi: 10.1016/j.heares.2013.09.012

Morgan, C. L. (1891). *Animal Life and Intelligence*. Boston, MA: Ginn & Company.

Morgan, C. L. (1896). *Habitat and Instinct*. London: Edward Arnold Press.

Morgan, C. L. (1920). *Animal Behaviour*. London: Edward Arnold Press.

Mosing, M. A., Verweij, K. J. H., Madison, G., Pedersen, N. L., Zietsch, B. P., and Ullen, F. (2014). Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evol. Hum. Behav.* 36, 359–366. doi: 10.1016/j.evolhumbehav.2015.02.004

Nakata, T., and Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behav. Dev.* 27, 455–464. doi: 10.1016/j.infbeh.2004.03.002

Nettl, B. (2000). "An ethnomusicologist contemplates universals in musical sound and musical culture," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, UK: The MIT Press), 463–472.

Nikolsky, A. (2015). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: Prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.0140

Okanoya, K. (2013). "Finite-state song syntax in bengalese finches: sensorimotor evidence, developmental processes, and formal procedures for syntax extraction," in *Birdsong, Speech, and Language : Exploring the Evolution of Mind and Brain*, eds J. J. Bolhuis and M. Everaert (Cambridge, UK: MIT Press), 229–242.

Opacic, T., Stevens, C., and Tillmann, B. (2009). Unspoken knowledge: implicit learning of structured human dance movement. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1570–1577. doi: 10.1037/a0017244

Osborn, H. F. (1896). A mode of evolution requiring neither natural selection nor the inheritance of acquired characters. *Trans. N.Y. Acad. Sci.* 15, 141–142.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York, NY: Oxford University Press.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nn1082

Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford; New York, NY: Oxford University Press.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998). Processing syntactic relations in language and music: an event-related potential study. *J. Cogn. Neurosci.* 10, 717–733. doi: 10.1162/089892998563121

Pearce, E., Launay, J., and Dunbar, R. I. M. (2015). The ice-breaker effect : singing mediates fast social bonding. *R. Soc. Open Sci.* 2, 1–9. doi: 10.1098/rsos.150221

Pearce, E., Launay, J., van Duijn, M., Rotkirch, A., David-Barrett, T., and Dunbar, R. I. M. (2016). Singing together or apart: the effect of competitive and cooperative singing on social bonding within and between sub-groups of a university Fraternity. *Psychol. Music* 44, 1255–1273. doi: 10.1177/0305735616636208

Pearce, E., MacCarron, P., Launay, J., and Dunbar, R. I. M. (2017). Tuning in to others: exploring relational and collective bonding in singing and non-singing groups over time. *Psychol. Music* 45, 496–512. doi: 10.1177/0305735616667543

Perlovsky, L. (2010). Musical emotions: functions, origins, evolution. *Phys. Life Rev.* 7, 2–27. doi: 10.1016/j.plrev.2009.11.001

Perlovsky, L. (2012). Cognitive function, origin, and evolution of musical emotions. *Music. Sci.* 16, 185–199. doi: 10.1177/1029864912448327

Podlipniak, P. (2015). "The origin of music and the Baldwin effect," in *Proceedings of Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music,* eds J. Ginsborg, A. Lamont, and S. Bramley (Manchester: Royal Northern College of Music), 671–677.

Podlipniak, P. (2016). The evolutionary origin of pitch centre recognition. *Psychol. Music* 44, 527–543. doi: 10.1177/0305735615577249

Purves, D. (2017). *Music as Biology: The Tones We Like and Why*. Cambridge: Harvard University Press.

Reber, A. S., Allen, R., and Reber, P. J. (1999). "Implicit versus Explicit Learning," in *The Nature of Cognition*, ed. R. J. Sternberg (Cambridge: MIT Press), 475–514.

Remijsen, B. (2014). The study of tone in languages with a quantity contrast. *Lang. Doc. Conserv.* 8, 672–689. Available online at: http://hdl.handle.net/10125/24620

Remijsen, B., and Gilley, L. (2008). Why are three-level vowel length systems rare? Insights from Dinka (Luanyjang dialect). *J. Phon.* 36, 318–344. doi: 10.1016/j.wocn.2007.09.002

Repp, B. H. (2007). Hearing a melody in different ways: multistability of metrical interpretation, reflected in rate limits of sensorimotor synchronization. *Cognition* 102, 434–454. doi: 10.1016/j.cognition.2006.02.003

Reybrouck, M. (2013). From sound to music: an evolutionary approach to musical semantics. *Biosemiotics* 6, 585–606. doi: 10.1007/s12304-013-9192-6

Richerson, P. J., Boyd, R., and Henrich, J. (2010). Gene-culture coevolution in the age of genomics. *Proc. Natl. Acad. Sci. U.S.A.* 107(Suppl.), 8985–8992. doi: 10.1073/pnas.0914631107

Roederer, J. G. (1984). The search for a survival value of music. *Music Percept. Interdiscip. J.* 1, 350–356. doi: 10.2307/40285265

Salimpoor, V. N., van den Bosch, I., Kovacevic, N., McIntosh, A. R., Dagher, A., and Zatorre, R. J. (2013). Interactions between the nucleus accumbens and auditory cortices predict music reward value. *Science* 340, 216–219. doi: 10.1126/science.1231059

Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8987–8992. doi: 10.1073/pnas.1414495112

Seifert, U., Verschure, P. F. M. J., Arbib, M. A., Cohen, A. J., Fogassi, L., Fritz, T., et al. (2013). "Semantics of internal and external worlds," in *Language, Music, and the Brain* (Cambridge; London: The MIT Press), 203–230. doi: 10.7551/mitpress/9780262018104.003.0008

Shannon, R. V. (2016). Is birdsong more like speech or music? *Trends Cogn. Sci.* 20, 245–247. doi: 10.1016/j.tics.2016.02.004

Sievers, B., Polansky, L., Casey, M., and Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expressions of emotion. *Proc. Natl. Acad. Sci. U.S.A.* 110, 70–75. doi: 10.1073/pnas.1209023110

Simpson, G. G. (1953). The Baldwin Effect. *Evolution* 7, 110–117.

Slocombe, K. E., Townsend, S. W., and Zuberbühler, K. (2009). Wild chimpanzees (Pan troglodytes schweinfurthii) distinguish between different scream types: evidence from a playback study. *Anim. Cogn.* 12, 441–449. doi: 10.1007/s10071-008-0204-x

Smith, J. D., Nelson, D. G., Grohskopf, L. A., and Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition* 52, 23–54. doi: 10.1016/0010-0277(94)90003-5

Stainsby, T., and Cross, I. (2008). "The perception of pitch," in *The Oxford Handbook of Music Psychology*, eds. S. Hallam, I. Cross, and M. Thaut (Oxford; New York, NY: Oxford University Press), 47–58.

Steinbeis, N., Koelsch, S., and Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological, and neural responses. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380

Stevens, C. (2004). Cross-cultural studies of musical pitch and time. *Acoust. Sci. Technol.* 25, 433–438. doi: 10.1250/ast.25.433

Stevens, C. J. (2012). Music perception and cognition: a review of recent cross-cultural research. *Top. Cogn. Sci.* 4, 653–667. doi: 10.1111/j.1756-8765.2012.01215.x

Storr, A. (1992). *Music and the Mind*. New York, NY: Ballantine Books.

Strait, D. L., and Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hear. Res.* 308, 109–121. doi: 10.1016/j.heares.2013.08.004

Suzuki, T. N., Wheatcroft, D., and Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* 7:10986. doi: 10.1038/ncomms10986

Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., and Wilson, S. J. (2014). The genetic basis of music ability. *Front. Psychol.* 5:658. doi: 10.3389/fpsyg.2014.00658

Tarr, B., Launay, J., and Dunbar, R. I. M. (2014). Music and social bonding: "Self-other" merging and neurohormonal mechanisms. *Front. Psychol.* 5:1096. doi: 10.3389/fpsyg.2014.01096

Tettamanti, M., and Weniger, D. (2006). Broca's area: a supramodal hierarchical processor? *Cortex* 42, 491–494. doi: 10.1016/S0010-9452(08)70384-8

Tillmann, B. (2005). Implicit investigations of tonal knowledge in nonmusician listeners. *Ann. N.Y. Acad. Sci.* 1060, 100–110. doi: 10.1196/annals.1360.007

Tillmann, B., Bharucha, J. J., and Bigand, E. (2000). Implicit learning of tonality: a self-organizing approach. *Psychol. Rev.* 107, 885–913. doi: 10.1037/0033-295X.107.4.885

Tinbergen, N. (1963). On aims and methods of Ethology. *Z. Tierpsychol.* 20, 410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x

Toates, F. M. (1988). "Motivation and emotion from a biological perspective," in *Cognitive Perspectives on Emotion and Motivation*, eds V. Hamilton, G. H. Bower, and N. H. Frijda (Dordrecht: Springer), 3–35.

Van de Cavey, J., and Hartsuiker, R. J. (2016). Is there a domain-general cognitive structuring system? Evidence from structural priming across music, math, action descriptions, and language. *Cognition* 146, 172–184. doi: 10.1016/j.cognition.2015.09.013

van de Waal, E., Borgeaud, C., and Whiten, A. (2013). Potent social learning and conformity shape a wild primate's foraging decisions. *Science* 340, 483–485. doi: 10.1126/science.1232769

Waddington, C. H. (1953a). Genetic assimilation of an acquired character. *Evolution* 7, 118–126.

Waddington, C. H. (1953b). The "Baldwin effect", "Genetic Assimilation", and "Homeostasis." *Evolution* 7, 386–387.

Wakita, M. (2014). Broca's area processes the hierarchical organization of observed action. *Front. Hum. Neurosci.* 7:937. doi: 10.3389/fnhum.2013.00937

Winter, Y. (2014). On the grammar of a Senegalese drum language. *Language* 90, 644–668. doi: 10.1353/lan.2014.0061

Wong, A. W.-K., Huang, J., Chen, H.-C., Chen, H., and Hills, M. (2012). Phonological units in spoken word production: insights from cantonese. *PLoS ONE* 7:e48776. doi: 10.1371/journal.pone.0048776

Woolhouse, M., Cross, I., and Horton, T. (2016). Perception of nonadjacent tonic-key relationships. *Psychol. Music* 44, 802–815. doi: 10.1177/0305735615593409

Xu, L., Thompson, C. S., and Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J. Acoust. Soc. Am.* 117, 3255–3267. doi: 10.1121/1.1886405

Zahavi, A. (1975). Mate selection—A selection for a handicap. *J. Theor. Biol.* 53, 205–214. doi: 10.1016/0022-5193(75)90111-3

Zahavi, A., and Zahavi, A. (1997). *The Handicap Principle : A Missing Piece of Darwin's Puzzle*. New York, NY; Oxford: Oxford University Press.

Zatorre, R. J., and Baum, S. R. (2012). Musical melody and speech intonation: singing a different tune. *PLoS Biol.* 10:e1001372. doi: 10.1371/journal.pbio.1001372

Zimmermann, E., Leliveld, L., and Schehka, S. (2013). "Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices," in *Evolution of Emotional Communication: from Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford; New York, NY: Oxford University Press), 116–132.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Podlipniak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 30 October 2017
doi: 10.3389/fpsyg.2017.01894

# A Joint Prosodic Origin of Language and Music

*Steven Brown**

*Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON, Canada*

Vocal theories of the origin of language rarely make a case for the precursor functions that underlay the evolution of speech. The vocal expression of emotion is unquestionably the best candidate for such a precursor, although most evolutionary models of both language and speech ignore emotion and prosody altogether. I present here a model for a joint prosodic precursor of language and music in which ritualized group-level vocalizations served as the ancestral state. This precursor combined not only affective and intonational aspects of prosody, but also holistic and combinatorial mechanisms of phrase generation. From this common stage, there was a bifurcation to form language and music as separate, though homologous, specializations. This separation of language and music was accompanied by their (re)unification in songs with words.

**Keywords: language, music, speech, song, evolution, prosody, intonation, emotion**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky,*
*Independent Researcher,*
*United States*

***Reviewed by:***
*Reyna L. Gordon,*
*Vanderbilt University, United States*
*Elizabeth Hellmuth Margulis,*
*University of Arkansas, United States*

***Correspondence:**
*Steven Brown*
*stebro@mcmaster.ca*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *19 June 2017*
***Accepted:*** *12 October 2017*
***Published:*** *30 October 2017*

***Citation:***
*Brown S (2017) A Joint Prosodic Origin of Language and Music. Front. Psychol. 8:1894. doi: 10.3389/fpsyg.2017.01894*

Theories of the origins of language generally fall into the two broad categories of vocal and gestural models (Corballis, 2002; MacNeilage and Davis, 2005; Armstrong and Wilcox, 2007; Arbib, 2012; McGinn, 2015). Given that humans have evolved species-specific capacities for both vocal imitation and gestural imitation (Donald, 1991), a central question is whether language evolved initially as a system of vocalization or one of gesture, since imitative mechanisms are critical to evolutionary accounts of language acquisition. Gestural theories of language have grown in popularity in recent years due to their association with mirror-neuron-based models of action observation (Arbib, 2012). However, vocal theories have a far deeper grounding in historical models of language, going back to the ancient Greeks. During the Renaissance, not only was it commonplace to talk about the evolutionary connection between language and music, but both functions were seen as being clearly rooted in the vocal expression of emotion (Condillac, 1746; Rousseau, 1781; Thomas, 1995), a trend that continued into Darwin's day (Spencer, 1857, 1890; Darwin, 1871, 1872) and through to the early twentieth century (Wallaschek, 1891; Newman, 1905; Nadel, 1930; Sachs, 1943).

While contemporary vocal accounts of language origin do not deny the linkage between speech and emotion, they do not consider it to be central to their models, focusing instead on the articulatory innovations of speech—such as complex phonemic repertoires, the nature of syllable structure, vocal learning, descent of the human larynx, among others—or the origins of symbolization *per se*, separate from emotional communication. Some models talk about the origins of speech in "singing" (Darwin, 1871; Jespersen, 1922), but there are problems associated with this invocation of singing. Singing as a human behavior implies something musical, but the musicality of the posited ancestral singing mechanism is not specified. Singing simply becomes a counter-state to speaking (i.e., language-based vocalizing), rather than being something truly musical, as predicated on the tonal principles of scale structure. When Jespersen (1922) claimed that our ancestors "sang out their feelings long before they were able to speak their thoughts" (p. 436), his notion of singing included such diverse vocalizations as the singing of birds, the roaring of

mammals, and the crying of babies. Likewise, Fitch (2010) referred to music as being an example of "bare phonology," viewing music as basically a counter-state to propositional speech. The aim of this essay is to propose a joint prosodic model of the origins of language and music, but to avoid the pitfalls of talking about a singing or phonological mechanism that has no musical specifications. As with my earlier writings on the topic (Brown, 2000a,b, 2007), my focus here will be on phylogenetic issues of cognitive structure, rather than on Darwinian issues of adaptiveness and selection mechanisms (which I have discussed in detail in Brown, 2000b).

While vocal and gestural models have generally been placed in opposition to one another, it is far more reasonable instead to see vocalization and gesture as complementary communicative specializations (McNeill, 2005; Arboitiz and Garcia, 2009; Arbib, 2012; Arboitiz, 2012; Garcia et al., 2014), as suggested in **Figure 1**. Gesture seems particularly well-suited to iconically convey information about the spatial properties of objects and actions through pantomimic gestures (Armstrong and Wilcox, 2007), which the vocal system cannot easily achieve through iconic means. By contrast, vocal prosody seems better suited to convey information about the emotional meaning of a perceived object for the communicator, in other words its consequentiality. To my mind, the co-speech gestures of modern speech are essentially pantomimes (Beattie, 2016), and might therefore comprise "fossils" of an early gestural stage of language evolution that was pantomimic. Another potential fossil consists of what I will call "acoustic pantomimes," namely iconic sounds, such as onomatopoeias. Such pantomimes are able to represent the sound-generating properties of objects and actions—as in the "ruff" of a dog barking—as well as non-vocal object-properties like size, height, velocity, and temperature that are correlated with the acoustics of objects (Nygaard et al., 2009; Dingemanse et al., 2015, 2016; Perlman et al., 2015; Svantesson, 2017). **Figure 1** suggests that pantomimes in both the vocal and gestural domains served as parallel precursor stages on the road to symbolization for each route of communication.

Beyond the intermediate iconic stage of pantomime, the bulk of the symbolic function of language resides with acoustic symbols, rather than gestural symbols, with the prominent exception of sign languages among the deaf. While I am quite sympathetic to gestural models of language origin, I will put them aside from this point onward in order to examine the basic question raised above, namely whether language and speech arose as an offshoot of the vocal expression of emotion. Most evolutionary speech models ignore prosody altogether and instead focus on the anatomical changes to the human articulatory system that permitted the emergence of syllable structure, including descent of the larynx. An interesting example is MacNeilage's (1998) "frame/content" model, which proposes that mandibular oscillations in great apes (e.g., lip smacking) provided a scaffold upon which syllable structure in speech may have arisen (see also MacNeilage and Davis, 2005; Ghazanfar et al., 2012). What is central to this model is that such oscillations are voiceless in non-human primates, and that the key innovation for speech would be the addition of phonation to such oscillations so as to create alternations between vowels (open vocal tract) and consonants (closed or obstructed vocal tract).

A central tenet of syllabic accounts of language evolution is the notion of "duality of patterning" (Hockett, 1960; Ladd, 2012), which argues that the acoustic symbols of speech—i.e., words—are built up from meaningless constituents. Words are comprised of fundamental units called phonemes, but none of the phonemes themselves have intrinsic meanings (although sound-symbolic accounts of language origin argue that phonemes have non-random occurrences across word classes and hence may have some minor symbolic content; Monaghan et al., 2016). Through the kinds of mandibular-oscillatory mechanisms that MacNeilage's (1998) speech model highlights, phonemes get combined to form syllables, most universally as alternations between consonants and vowels. This process of syllable formation is not merely oscillatory but is *combinatorial* as well. From a small and fixed inventory of consonants and vowels—generally a few dozen in any given language—these phonemes become combined to form syllables. Such syllables may constitute words in and of themselves ("bee"), or they may be combined with other syllables to form polysyllabic words ("being," "Beatrice"). Finally, through a different type of combinatorial mechanism, words can be combined with one another to form phrases and sentences through compositional syntactic mechanisms, as described below in the section "Syntax evolution and the 'prosodic scaffold' ".

## THE PROSODIC SCAFFOLD

I would like to reflect on what is missing in the standard syllabic account of speech and language just presented. Much of it comes down to whether one thinks of language evolution as serving a purely cognitive function for an individual (Berwick, 2011) or instead a communicative function for a group of individuals (Tomasello, 2003; Robinson, 2005; Scott-Phillips, 2017). In a later section of this article about syntax, I will describe this as a contrast between a "monologic" view (focused on internal thought) and a "dialogic" view (focused on social interaction) of language evolution. If one thinks about language and speech as a dialogic system of social communication, as many theorists suggest, then vocal prosody is critically missing from the syllabic account.

Prosody is characterized by a number of expressive melodic and rhythmic features of an utterance that convey information about emotion, intention, attentional focus, and communicative stance (Scherer, 2003). It is quite different from what has been described for syllables. It is not combinatorial, but is instead holistic, conveying emotional meanings according to rules of expression that govern emotional modulations of vocal pitch, loudness, timing/duration, and timbral features, as based on the valence and intensity of the expressed message. An influential cross-species account of this is found in Morton's (1977) set of "motivation-structure rules," themselves based on thinking going back to Darwin's (1872) treatise on *The Expression of Emotions in Man and Animals*. For example, aggression is conveyed with harsh, low-frequency sounds, whereas fear and submission are conveyed with more tone-like, high-frequency

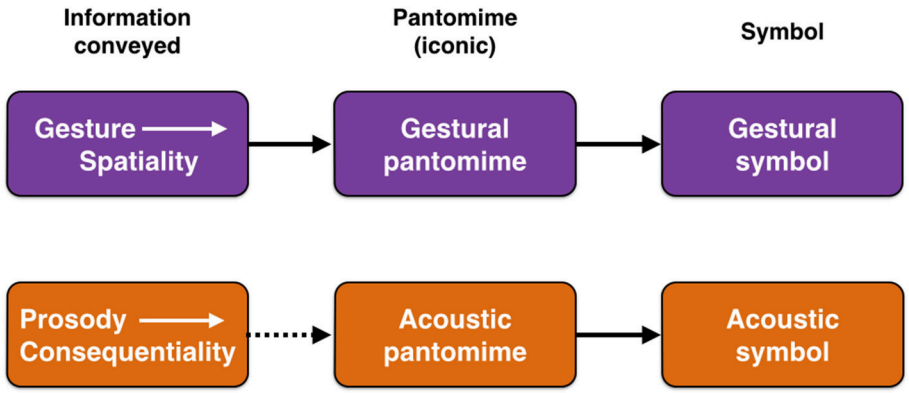



**FIGURE 1 |** Prosody and gesture compared. The figure shows parallel gestural and vocal routes to the origin of language by means of either gestural symbols (sign language) or acoustic symbols (speech). The pantomime stage in the middle of both routes is one of iconic representation. The term "acoustic pantomime" refers to iconic words (onomatopoieas) and so-called sound symbolisms. The dashed arrow in the second row suggests that, compared to the more natural relationship between gesture's spatiality and visual pantomimes, the connection between prosody's conveyance of consequentiality and acoustic pantomimes is more remote.

sounds. According to a prosodic account, it is not sufficient to think of "bee" as an arbitrary acoustic symbol for an insect that is generated through the combination of meaningless phonemes. In actual interpersonal communication, "bee" will be vocalized in such a manner as to convey the consequentiality of that insect for the speaker and his/her audience, as governed by prosodic rules of expression that communicate the emotional valence and intensity of that insect for the communicators. In other words, the vocal expression "Bee!!" during interpersonal communication conveys as much about the speaker's emotions and intentions as it does about the object itself. The holistic nature of prosody operates such that the declamation "Bee!" constitutes a complete utterance; it is essentially a one-word sentence.

It is important to consider that prosody is not some add-on to the combinatorial phonemic mechanism for generating syllable strings and sentences, but instead the reverse: it is the foundation of vocal communication, not least speech. Phonemic mechanisms must be superimposed upon *it*; it is not the case that a monotone string of phonemes becomes "melodized" by prosody after the fact. Prosody is intrinsic to the generative mechanism of speech and is in fact the primary consideration in communication, namely the speaker's emotional intent and message. While there is ample evidence for a "prosody first" model of speech planning when applied to the linguistic-prosodic levels of phonological and phonetic encoding (Keating and Shattuck-Hufnagel, 2002; Krivokapic, 2007, 2012), there is still minimal experimental work regarding the generative aspect of the expression of affect in speech. Instead, most language models place "conceptual structure" at the highest level of the communicative hierarchy (Levelt, 1999), implicating the domain of semantics and thus words. What is missing here is an overarching "emotional semantics" of communication in which emotion is a primary component in the production of speech, preceding word selection. Other theorists have made similar claims with reference to "ostensive communication" or the communication of intent (Scott-Phillips, 2017). Consider the sentence "It's a bee." That same string of words would not only be uttered in a dramatically different manner between seeing a photograph of a bee in a magazine as compared to seeing an actual bee on one's dinner plate, but the behavioral consequences would, in theory, be quite different. So, while linguists are able to support a "prosody first" model when it comes to linguistic prosody vis-à-vis syntax, I would argue that we need to expand this to have the planning of affective prosody occur at an even earlier stage in the process.

Based on this reasoning, I would like to propose a "prosodic scaffold" model in which overall communicative (intentional, emotional) meaning is the primary factor being conveyed in speech and in which the combinatorial and compositional mechanisms of speech's words and utterances act to "fill out" a prosodic scaffold. This scaffold is comprised of (1) affective prosody, which refers to the vocal expression of emotions, usually acting on the utterance at a global level; and (2) linguistic prosody, which refers to a set of both local and global mechanisms for conveying emphasis (stress, prominence, focus), modality (e.g., question vs. statement), among other features (Cruttenden, 1997; Ladd, 2008). Because I am going to apply concepts about linguistic prosody to music in this article, I will avoid confusion in nomenclature by referring to it as "intonational" prosody from this point on.

## Affective Prosody

There are affective mechanisms that modulate the overall pitch height, loudness, and tempo of a spoken utterance in order to convey the emotional valence and intensity of the communicator's meaning. For any given sentence, happiness is typically conveyed by high, loud, and fast prosody, while sadness is conveyed by the opposite profile (Banse and Scherer, 1996). The same is true for the expression of these two emotions in music (Juslin and Laukka, 2003). These types of affective prosodies often work in a global fashion, affecting the entire scaffold of the phrase. For example, happiness both moves the scaffold to a higher vocal register and compresses it in time,

while sadness moves it to a lower register and expands it in time. In both cases, the holistic formula of a declarative sentence is preserved, but its acoustic properties are modified by emotional intent.

## Intonational Prosody

A majority of spoken utterances are declarative and are characterized by stereotyped intonations having either arched or descending pitch contours (Halliday, 1967; Cruttenden, 1997). This is exemplified in the top row of **Figure 2** for a single sentence using musical notation, which is taken from Chow and Brown's (under review) analysis of the relative-pitch profiles of spoken sentences, as averaged across a group of 19 speakers. One can readily observe the basic pattern of declination (i.e., falling pitch) characteristic of declarative sentences. The holistic nature of this formula is shown by the fact that, when the sentence is lengthened through the addition of words at the end, the declination process is *suspended* until a later point in the sentence. What this indicates is that the holistic prosodic scaffold of a declarative sentence, with its descending contour, is "filled out" in the longer sentence by suspending the point of declination until the terminal word.

If we contend that the vocal expression of emotion was the precursor to speech, then the evolution of the phonemic combinatorial mechanism had to find a way to create words (strings of segmental units) and phrases in the context of communicating emotional meanings by filling out a prosodic scaffold. The alternative idea, namely that prosody is some type of *post-hoc* affective modulation of a previously-established linguistic message, seems untenable and should not serve as the basis for evolutionary models of language and speech. While there are clearly non-prosodic means of conveying emotion and intention in speech, such as through word selection and syntactic constructions, these do not circumvent the need to be externalized through a prosodic-scaffold mechanism during vocal communication.

Before moving on to present my evolutionary model of a prosodic precursor to speech and music, I will summarize the model briefly so as to facilitate the presentation (which is outlined in **Figure 6** below, as described in the section "Bifurcation to form language and music"). I will argue that there was not one but two shared stages that served as joint precursors in the evolution of language and music: (1) the first was a system of affective prosody, and (2) the second was a system of intonational prosody. In other words, affective and intonational prosodies evolved through a sequential process as two linked evolutionary stages. In addition, while the first stage was made up of innate calls, the second capitalized on the newly-evolved capacity for vocal learning in humans. Following this second joint stage, language and music branched off as reciprocal specializations, each one retaining certain key features of their joint precursor stages. The model attempts to account for modern-day similarities between music and language/speech by loading the precursor stages with as many shared features as is theoretically justified. I argued in Brown (2000a) that, given that language and music possess both shared and distinct features, it would be most parsimonious to propose that their shared features evolved first, and that their domain-specific features evolved later as part of a branching process (see also Mithen, 2005), making language and music homologous functions (Brown, 2001). This idea would stand in contrast to models contending that music evolved from speech (Spencer, 1857), that speech evolved from music (Darwin, 1871; Jespersen, 1922; Fitch, 2010), or that music and language's similarities arose independently by convergent evolution.

Before proceeding to describe the model, I want to point out that, given the absence of any clear definition of music, I am going to adopt a view of music (and singing) that leans heavily on the side of pitch and most especially on the use of scaled pitches, even if there is imprecision in the scale degrees and/or their execution by a voice or instrument (Nikolsky, 2015). As a result, I am going to distinguish music from both speech prosody and emotive vocalizations. In addition, while rhythm is a critical feature of music, there is no aspect of music's rhythmic system that is not potentially shared with either dance or metric forms of speech, like poetry. Hence, if the development of an evolutionary model of music requires that we identify domain specificity for music, then I see tonality as the principal distinguishing feature of music (see Savage et al., 2015 for ethnographic support for this). While I am familiar with myriad examples of musics that fail this definition—e.g., they are based on unpitched percussion sounds, they are pitched but are not based on fixed scales, they are more concerned with timbral changes than pitch pitches, they contain emotive vocalizations, prosodic speech, and/or whispers—I cannot see the utility of developing an evolutionary account of music based on either non-tonal or metrical features. Instead, my model posits that a non-tonal prosodic precursor was the source for the tonal system that came to characterize much music as we know it.

The first evolutionary stage of the prosodic model of language origin is proposed to be a system of affective calling derived from the mechanisms of emotional vocalizing found in mammals more generally (Briefer, 2012). I have argued previously (Brown, 2007) that, not only was this particular affective system a joint precursor for language and music, but that it was a *group* communication system, particularly one that operated in ritualized contexts, such as territory maintenance, and that acquired its group force though emotional contagion (see also Hagen and Bryant, 2003). Using a wolf chorus as a model of such a precursor, I argued that this evolutionary stage was characterized by an imprecise overlapping of parts among the group members, showing little to no synchronization of parts. This performance arrangement is referred to musically as *heterophony*[1], which is when "different parts are performing the same melody at the same time but with different versions" (Malm,

[1]I am using the term heterophony in this article so as to be consistent with similar ideas about the evolution of musical texture that I described in Brown (2007) in an article entitled "contagious heterophony". However, the editor of this research topic, Aleksey Nikolsky, takes issue with my usage, and suggests that his own term "isophony" (Nikolsky, 2016, Appendix V) is a better description of the choral texture that I am alluding to. He defines isophony as a "brief call, continuously reproduced by multiple performers with irrational deviations in timing and pitch, where each participant retains idiosyncrasy of the rhythmic, timbral, and directional attributes of the pitch contour – altogether producing a 'jumbled' effect." This strikes me as an excellent description of the phenomena that I am referencing here. Readers are encouraged to see Nikolsky's commentary on the current article for a more detailed description of isophony as applied to the evolutionary origins of musical/linguistic communication in humans.

1996 p.15). Starting from this common, asynchronous precursor, an evolutionary branching process would occur to create two different forms of coordination during communication such that (1) music would evolve to achieve a tight temporal *integration* of parts through the evolution of the capacities for both rhythmic entrainment and vocal imitation, and that (2) speech would evolve to achieve an *alternation* of parts, as occurs in standard dialogic exchange (**Figure 3**). This functional and structural bifurcation reflects the fact that music retains the precursor's primary function as a device for group-wide coordination of emotional expression, most likely for territorial purposes, while language evolves as a system for dyadic information exchange in which an alternation of parts, rather than simultaneity, becomes the most efficient means of carrying out this exchange. These distinctive communicative arrangements of music and speech come together in a performance style that is found quite widely across world cultures, namely *call-and-response* singing (**Figure 3**), where the call part is informational and is textually complex (typically performed by a soloist, as in speech) and the response part is integrated and textually simple (typically performed by a chorus, as in music). Call-and-response is an alternating (turn-taking) exchange, but one between an individual and a group, rather than two individuals.

The idea that speech evolved from a group-wide communication system has a distinct advantage over individualist accounts of language origin in that can provide a solution to the "parity" problem (Arbib, 2012). The evolution of communication systems is constrained by the fact that meanings have to be mutually interpretable in order to be adopted by a community of users. Any person can devise a word to mean bee, but unless everyone in the community understands it, then it is useless as anything more than an unintelligible device for self-expression. A group-communication system obviates this problem, since it is produced collectively. In addition, and following along the lines of the wolf example, making the group-communication system something ritualized helps in achieving meaning through the use of context specificity and the signaling of consequentiality. Communication will occur in situations that have shared emotional meanings and shared consequences for all members of the group, such as during territory defense. Having language be group-level from the start provides a solution to a number of evolutionary obstacles to achieving parity in communication.

## "MUSILANGUAGE" AS A JOINT PROSODIC PRECURSOR

This first precursor stage of group-affective vocalizations that I have just described would be a ritualized territorial chorus of emotional communication. It would be neither speech-like nor music-like in its acoustic features, but instead something similar, functionally and structurally, to a non-human form of group chorusing, like a wolf chorus or a pant hoot chorus in chimpanzees. While speech and music do indeed have shared

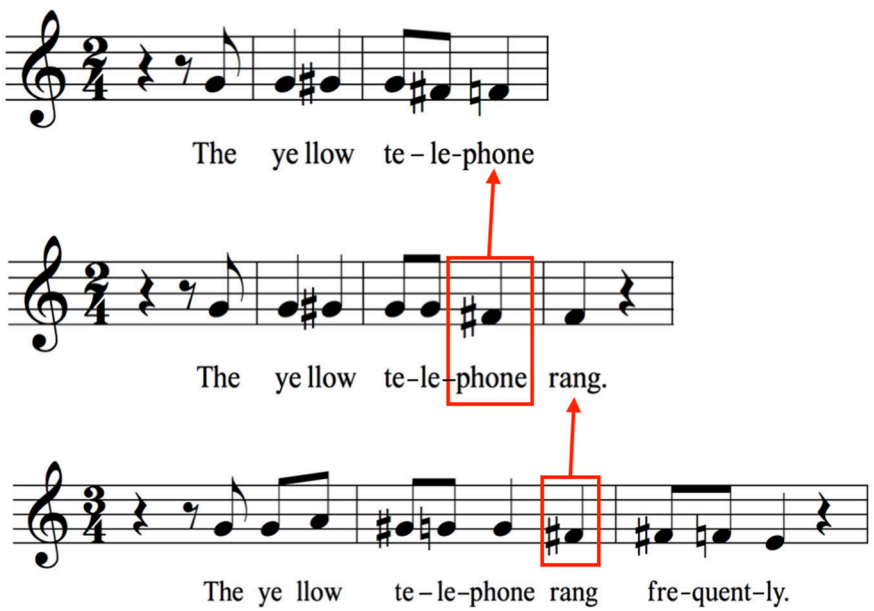



**FIGURE 2 |** Prosodic scaffolds. The notion of a prosodic scaffold refers to the melodorhythmic shell of a spoken utterance, as based on its prosodic features. Many declarative sentences have simple declining pitch profiles, as shown in musical notation on the top line for the phrase "The yellow telephone." A lengthening of the utterance (middle line) produces a suspension of the declination found in the shorter phrase, as shown by the red box around "phone," which is higher in pitch than "phone" in the first phrase. The same mechanism occurs again on the bottom line when the utterance is lengthened a second time. Here the suspension occurs on "rang" (red box), which is higher than "rang" in both preceding sentences. Data are taken from Chow and Brown (under review), based on the relative-pitch productions of 19 native speakers of English. The relative pitch is shown on a treble clef for convenience of visualization. The absolute pitches are about an octave lower.

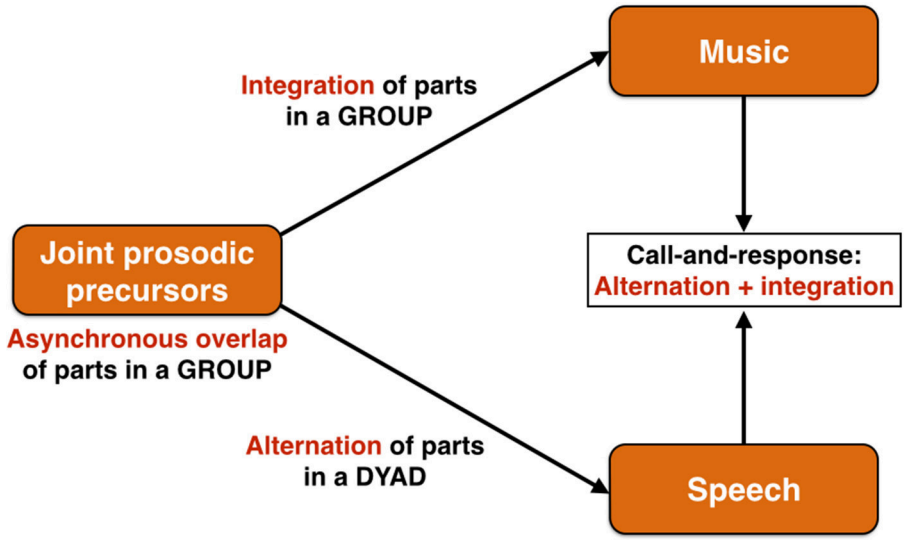



**FIGURE 3 |** The performance arrangements of communication. The joint prosodic precursor is proposed to be a group communication system based on both affective and intonational prosodies and that is characterized by an asynchronous overlapping of parts among the individual callers. An evolutionary branching process occurs such that music would evolve to achieve a tight temporal integration of parts in a group, whereas speech would evolve to achieve an alternation of parts in a dyad, as occurs in standard conversational exchange. These distinctive communicative arrangements of music and speech are joined together in call-and-response singing, where the call part is generally performed by a soloist in alternation with a chorus. It is essentially an alternation between an individual and a group, thereby showing a combined profile of music's and speech's distinctive performance arrangements.

mechanisms of emotional expression (affective prosody), the affective precursor just described is lacking in many additional features that are shared between speech and music and that should be reasonably found in a joint evolutionary precursor. Hence, my model requires the existence of a second joint precursor-stage before the bifurcation occurred to generate language and music as distinct and reciprocal specializations emanating from it. While the first stage focused on the shared features of affective prosody, this second stage should now contain the shared features of *intonational* prosody that are found in speech and music.

My characterization of this second precursor stage will comprise a revised and corrected account of what I called the "musilanguage" system in a previous publication (Brown, 2000a) and which was fleshed out in book form by Mithen (2005). Hence, it will comprise my Musilanguage 2.0 model. The core idea of the model is that *those features of language and music that are shared evolved before their domain-specific features did* due to the presence of a joint precursor—what I call the musilanguage system—that contained those shared features. While the initial joint precursor described above would be a system of affective prosody, this second stage would achieve the next important level of intonational prosody. In other words, *it would embody those features of intonation that are shared between language and music*, but without requiring lexicality, syntax, or musical scale structure (tonality), an idea also developed by Fitch (2010). If I had to summarize the properties of this stage, I would argue that it is a "grammelot," in other words a system of nonsense vocalizing or pure prosody, as was prominent as a tool for traveling theater companies in the days of the Commedia dell'Arte during the Renaissance (Jaffe-Berg, 2001; Rudlin, 2015). The only modification that I would propose is that the musilanguage precursor was a *group-level* grammelot, produced through chorusing. In what follows, I will outline a number of key properties of the proposed joint prosodic precursor, with an eye toward defining those features that can be thought of as shared between speech and music and hence that can be most reasonably attributed to a joint evolutionary stage. **Table 1** lists a dozen such features.

## Voluntary Control of Vocalization and Vocal Learning

I contend that the transition from the first affective stage to this second stage corresponds to the transition from the non-human-primate system of involuntary control of stereotyped calls to the appearance in humans of both voluntary control over the vocal apparatus and vocal production learning. Belyk and Brown (2017) proposed a co-evolutionary account for the joint appearance of these two capacities, although other theorists have suggested sequential models (Ackermann et al., 2014). Hence, the advent of the musilanguage stage would mark a transition from innate to learned vocalizing, accompanied by the complexification of communication sounds that learning makes possible. This similarity between speech and music as learned vocal-communication systems is an extremely important one to keep in mind. *It places the emergence of vocal learning firmly upstream of the separation between language and music in human evolution.*

**TABLE 1 |** Features of the musilanguage system.

| | |
|---|---|
| 1 | Vocal production learning |
| 2 | Breath phrases |
| 3 | Level tones and level transitions |
| 4 | Imprecise levels-and-contours pitch system (tonicity), with melodic motion based on pitch proximity |
| 5 | Phonemic combinatoriality (but as meaningless vocables) |
| 6 | Meaningful melodies: Holistic intonational formulas |
| 7 | Phrase structure uniting combinatorial and holistic processes |
| 8 | Affective prosody (inherited from the first precursor stage) |
| 9 | 2- and 3-unit stress groupings (headedness) |
| 10 | Heterometric rhythms |
| 11 | Repetitive form |
| 12 | Polyphonic and heterophonic performance arrangements |

## Breath Phrases

Phrase structure in both speech and music approximates the length of a breath phrase (Pickett, 1999). This may seem like a trivial similarity between speech and music as communication systems, but it also makes them natural partners when it comes to setting words to music (discussed below). There were significant changes in the voluntary control of respiration during human evolution (MacLarnon and Hewitt, 1999, 2004; Belyk and Brown, 2017), and it would seem that such changes impacted speech and music in comparable manners to influence the structural features of phrases in both domains. When people take breaths while either speaking or singing, they tend to do so at phrase boundaries, rather than in the middle of a phrase (Grosjean and Collins, 1979). In addition, the depth and duration of an inhalation correlate with the length of a produced sentence (Fuchs et al., 2013). Finally, extensive work on the analysis of pause duration as a function of the length and/or syntactic complexity of sentences points to a role of respiratory planning in speech production (Krivokapic, 2012). Hence, there is clear motor planning for speech at the level of respiration. Provine (2017) proposed that the nature of human breathing, and thus vocalization, may be a direct product of the transition to bipedal locomotion.

## Level Tones and Level Transitions

An important feature of human vocalization that is virtually never mentioned in evolutionary accounts of speech or music is the fact that humans can produce level tones when vocalizing. Much of primate vocalizing is based on pitch glides, as can be heard in the pant hoot of chimpanzees and the great call of gibbons. While such glides are still present in human emotional vocalizations, such as in cries, both speech and music are based on transitions between relatively discrete tones. These tones are generally longer and more stable in music than they are in speech, but level tones seem to be present in speech to a large extent (Mertens, 2004; Patel, 2008; Chow and Brown, submitted). The defining feature of much music is not only that the transitions are level but that they are *scaled* and *recurrent* as well. Hence, instead of having an imprecise sequence of tones, the tones become digitized to create a small set of pitches that are used recurrently across the notes of a melody, in both ascent and descent. When this does not occur, we get a melodic pattern that is speech-like (Chow and Brown, submitted), although such a pitch pattern sounds increasingly music-like the more it is tandemly repeated (Deutsch et al., 2011).

## Levels-and-Contours (L&C)

Building on the last point, a related acoustic feature of the musilanguage system is that it would be based on an imprecise (non-recurrent) and coarse-grained mechanism of pitch signaling that involved a basic sense of both pitch levels (e.g., high vs. low) and pitch contours (e.g., rising vs. falling). Importantly, this system would be *pitched but not tonal*. In other words, it would not be based on the scaled pitches that are found in the majority of musical systems, and would thus not be, in my view, a true form of singing. This idea is a modification of an incorrect proposal that I made in the original publication based on a limited database at the time about the pitch properties of speech (Brown, 2000a), about which Fitch (2010) was quite justified in raising objections. Instead, I now see the precursor system as having an imprecise relative-pitch system based on optimizing the contrast between relatively high/rising and relatively low/falling pitches, what I will refer to as a levels-and-contours (L&C) system. In twentieth century British theorizing about intonation, the term "tonicity" was used to characterize this type of pitch system (Halliday, 1967, 1970), where different types of pitch contours are used to signal intonational meaning. A key acoustic feature of this system that is shared between speech and music is that melodic movement tends to be based on pitch proximity, rather than large leaps (Huron, 2006; Patel, 2008; Chow and Brown, under review). I will propose below that, after the separation of language and music, *speech retained the imprecise levels-and-contours system of the precursor*, while music increased the precision of the pitch relationships by introducing tonality through scale structure, making the pitches recurrent in the formation of melodies and thereby making music into a combinatorial system for pitch. Hence, the coarse-grained pitch production mechanism of the levels-and-contours system of the musilanguage stage provides a reasonable joint precursor for the pitch properties of both speech and music.

## Phonemic Combinatoriality

One of my major contentions is that the evolution of phonemic combinatoriality is a feature that should be placed upstream of the divide between speech and music, comprising a key property of the joint musilanguage precursor (see **Figure 6** below). This conforms with Fitch's (2010) claim that proto-language was a system of "bare phonology." Tallerman (2013) takes issue with the concept of "phonology" being applied to anything other than meaningful words and thus true language, although I would point out that proto-language models do not present any kind of specification of the phonetic properties of their proto-words (Bickerton, 1995; Jackendoff, 1999). Hence, there was most likely a proto-phonology in place before language evolved. I mentioned MacNeilage's (1998) frame/content theory above, which seems to be as good a model as any for the origin of syllable

structure through phonemic combinatoriality. Most mandibular oscillations in non-human primates are voiceless, and so a critical feature of the MacNeilage model is that the open vocal-tract configuration of the oscillation should become phonated, making syllables into pitch-bearing units. As per the point raised in the previous two paragraphs, this should permit the formation of level tones as well as glides. As such, this would favor the use of open syllables at this stage, so as to maximize information due to pitch variation. Importantly, phonemic combinatoriality would provide one mechanism for creating phrase structure by the musilanguage system, such that the vocalic part of the syllable would serve as the locus of melodic and rhythmic variation. I could imagine that the musilangauge system of proto-phonology was comprised of a repertoire of such syllabic units (see **Figure 4**). Given that this stage preceded the evolution of lexicality, then these syllables were *vocables*, or nonsense syllables, in keeping with the musilanguage's status as a grammelot. As in many forms of birdsong, there could have been a large diversity of such units, even though each unit would be devoid of intrinsic meaning (Marler, 2000; Slater, 2000).

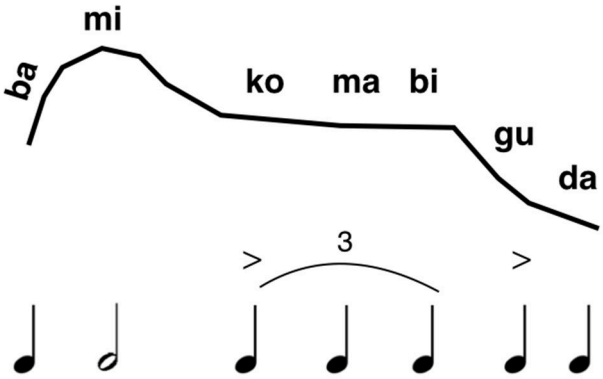



**FIGURE 4 |** A speculative proposal for a musilinguistic phrase. The figure presents a speculative example of what a musilinguistic phrase might be like. The top row indicates the relative-pitch profile of the melody, while the bottom row indicates the relative duration values of the syllables. Phonetically, the system is comprised of a repertoire of open syllables based on phonemic combinatoriality. Open syllables are shown here, with the pairing between consonant and vowel being based on MacNeilage's (1998) model of speech evolution.

## Meaningful Melodies: Holistic Intonational Formulas

Beyond the localist mechanism underlying phonemic combinatoriality, there would be a more global and holistic system of pragmatic intonational melodies that had categorical meanings (Fernald, 1989; Papousek, 1996), just as they do in modern speech. The most basic contrast would be between 1) phrases with descending contours that end in low tones, as in typical declaratives, conveying a sense of certainty, stability, and/or finality, and 2) phrases that proceed and/or end in high tones, conveying a sense of uncertainty, continuity, suspense, or incredulity (Halliday, 1967, 1970). The latter are perhaps the first questions of human communication (Jordania, 2006), conveyed via intonation alone through a grammelot, much the way that filtered speech retains the intonational meanings of sentences (Fernald, 1989). It would be hard to estimate how many melodies would exist in the repertoire of this system, but these would be pragmatically-distinct melodies that operated more or less in a categorical manner. Hence, this could be a first step in achieving phrase-level meanings and prosodic scaffolds before lexicality was introduced (Fitch, 2010), in which case the system would be better characterized as one of pure *prosody* than pure phonology. The resultant phrases could be thought of as "holophrases." However, these would not be the *symbolic* holophrases discussed by people like Wray (1998), Mithen (2005), and Arbib (2012), but instead *prosodic* holophrases that conveyed affective and pragmatic meanings in a holistic manner, much the way that speech prosody often does. This relates to some models of speech and/or music evolution that posit a central role for mother/infant communication (Dissanayake, 2000; Falk, 2009).

## Affective Prosody

By inheriting the innate expressive mechanisms from the first evolutionary stage—something that itself is phylogenetically derived from primate communication—the musilanguage system would have additional expressive modulation of phrases according to the valence and intensity of the communicated emotion, providing yet another influence on the melody and rhythm of the phrases. This occurs with regard to global and local changes in the pitch (register), loudness, and tempo of the phrases. I called these "expressive phrasing" mechanisms in the original publication (Brown, 2000a) and now see the initial affective precursor as being a specialized version of affective prosody occurring as a group territorial chorus.

## Stress Groups

The system should show similarities to features of stress timing seen in speech and music, whereby syllabic units often occur in 2- or 3-unit groupings, with a sense of stress on the initial syllable (Brown et al., 2017). This conveys what linguists call "headedness" (Jackendoff, 2011), which is a hierarchical differentiation of the elements within a grouping, where emphasis is generally placed on the first element. These groupings can themselves be organized hierarchically and can potentially be embedded in one another in a recursive fashion (Jackendoff, 2011; Tallerman, 2015). The musilanguage system is thus proposed to have hierarchical phrase organization, an important feature shared between music and speech (Lerdahl and Jackendoff, 1983; Lerdahl, 2001).

## Heterometric Rhythms

The rhythmic properties of this system would not be the isometric type of rhythm found in much music, but instead the "heterometric" type of rhythm that is characteristic of speech (Brown et al., 2017). Instead of having a single, fixed meter, the rhythm might involve changes in stress patterns, but still maintaining the primacy of 2- and 3-unit groupings and patterns.

## Repetitive Form

To the extent that such a communication system would be both ritualized and performed in groups, it might have a strongly

repetitive type of form (a so-called "ostinato" form), as in much music in indigenous cultures (Lomax, 1968) and beyond (Margulis, 2013, 2014). Hence, the same phrases would be uttered repeatedly. **Figure 4** presents a highly speculative account of what a musilinguistic phrase might look like in terms of (1) phonemic combinations to diversify the number of syllable types, (2) the predominant use of open syllables, (3) the overall melodic contour of an arching intonational formula (as one example of such a formula), and 4) the local grouping-structure of the rhythmic units, but with a non-metric rhythm overall. Such a phrase might be uttered repeatedly by a given individual during group chorusing. Compared to the vocalizations of the first evolutionary stage, this would be a learning-based system that permitted voluntary control of vocalizing, although still occurring in a ritualized manner at the group level.

## Polyphonic Texture

Finally, in order to think about the performance arrangement of the musilanguage system, **Figure 5** presents an overview of the major types of "textures" (i.e., multi-part performance arrangements) found in both human and animal chorusing. The figure is organized as a 2 × 2 scheme, where one dimension relates to pitch (whether the melodic lines are the same or not) and the other dimension relates to rhythm (whether the various parts are either synchronized in time or not). I argued in Brown (2007) that the initial evolutionary stage of group-level affective prosody was characterized by a "heterophonic" texture in which each individual of the group performed a similar melodic line but in which the parts were asynchronous in onset, as seen in a wolf chorus (see also **Figure 3** above). There are many examples of such chorusing in animals and humans (Filippi, 2016). In order to make the musilanguage stage more language-relevant, I would argue that, in addition to the presence of heterophony, this system would show the new texture of *polyphony*. Polyphony allows for two significant changes in the structural properties of performance compared to a heterophonic system: (1) there is a diversification of the vocal parts and hence the possibility of differentiation according to communicative roles, and (2) there is some degree of alternation of parts. The musilanguage stage would start to show some signs of alternation, which is a defining feature of conversation and a key feature of call-and-response musical forms. This is a first step toward having a differentiation of parts, both in terms of content and presentation, hence permitting a leader/follower dynamic. However, instead of having the seamless separation of parts that occurs in conversation, this stage would most likely have an imprecise type of exchange, in which the alternating parts overlapped with one another, as seen in a number of primate and avian duets, for example in gibbons and duetting birds (Dahlin and Benedict, 2013). One implication of the proposal that I am making here is that the capacity for vocal learning arose before the capacity for rhythmic entrainment and integration (contra the proposal of Patel, 2014). Hence, the musilanguage system, while voluntary and learned, would still have a relatively poor capacity for the synchronization of parts. I will return to this important point in the next section about the evolutionary changes that made music possible.

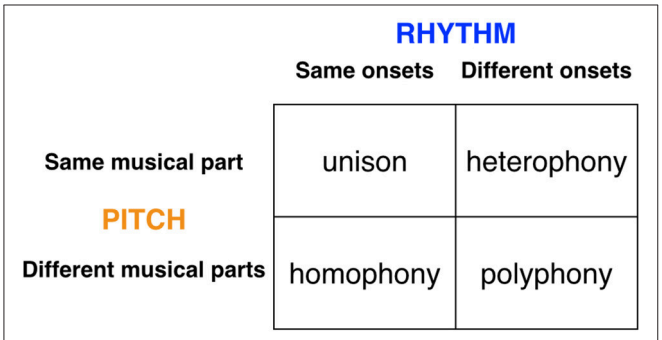



**FIGURE 5 |** Textures of chorusing. The figure presents an overview of the major types of textures (i.e., multi-part performance arrangements) found in chorusing, both animal and human. The figure is organized as a 2 × 2 scheme, where one dimension relates to pitch (whether the melodic lines are the same or not) and the other relates to rhythm (whether the various parts are synchronized in time or not). Each cell indicates a principal type of choral texture. The right-side cells are found in both animals and humans, while the left-side cells are principally found in humans. It is proposed that the initial affective precursor of language and music was heterophonic, while the musilanguage stage was potentially polyphonic as well. The aligned textures of unison and homophony required the emergence of the human capacity for metric entrainment in order to evolve.

What would be the function of the musilanguage system as a group-level grammelot compared to the first stage of innate affective expression? In keeping with my attempt to optimize the shared prosodic features of language and music before their separation, I would say that the system could be involved in group communication but in functions more dyadic as well. For example, a simple call-and-response interaction might be a novel arrangement of this system, showing some basic capacity for the alternation of parts and thus the roots of the information exchange that occurs in dialogue. While the syllabic units would be meaningless, the intonational melodies might be able to be used referentially to convey emotional meanings about objects in the environment or the actions of others, hence communicating consequentiality in a non-lexical and prosodic fashion.

# BIFURCATION TO FORM LANGUAGE AND MUSIC

With this description in mind of two sequential precursor-stages shared by language and music, we can now examine the bifurcation process to form full-fledged language and music as distinct, though homologous, functions, as well as their (re)unification in the form of songs with words, including call-and-response chorusing. **Figure 6** presents an overview of the model, starting with the innate group calling of affective prosody, followed by the musilanguage system of intonational prosody. The figure highlights the important proposal that phonemic combinatoriality is a shared feature of language and music, and this forms a critical part of what will be jointly carried over during the bifurcation process. I will first talk about language (lower part

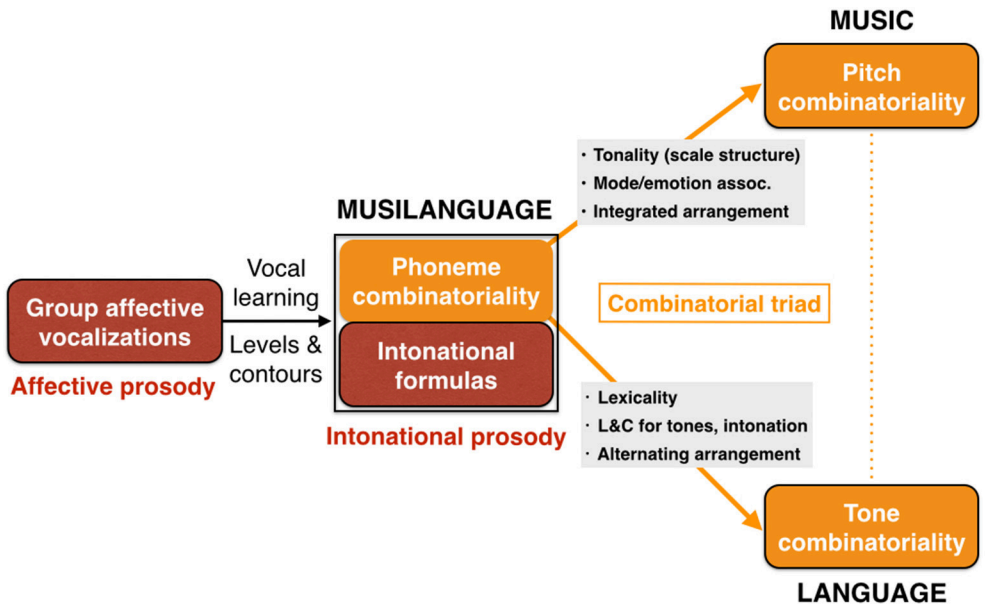



**FIGURE 6 |** The Musilanguage 2.0 model. Three evolutionary stages are shown. Stage 1 at left is one of group emotional vocalizations based on innate calls driven by the mechanisms of affective prosody. Stage 2 in the middle is the musilanguage stage of intonational prosody, possessing the various features outlined in **Table 1**, including vocal learning and the levels-and-contours pitch system. Two of its key features are highlighted here, first a system of phonemic combinatoriality (where orange in the figure signifies combinatorial), and second a system of holistic intonational formulas (where dark red signifies prosodic and holistic). Stage 3 at right is the bifurcation to form music and language as separate, though homologous, functions. The road to music involves a digitization of the pitch properties of the levels-and-contours precursor to develop tonality based on scale structure. This is accompanied by a system of emotional-valence coding that I call "scale/emotion" associations. The performance arrangement is integrated due to evolutionary changes permitting entrainment using metric rhythms. A domain-specific combinatorial feature of music is pitch combinatoriality. Next, the road to language involves the capacity to generate words through acoustic symbols. Externalization of language through speech is proposed to retain the levels-and-contours (LandC) system used by the musilanguage precursor. I propose that speech evolved as a lexical-tonal system from its inception, one that worked based on combinatorial principles. Language develops a performance arrangement that is based on alternation. In orange is presented a "combinatorial triad" of phonemic combinatoriality (shared between music and language), pitch combinatoriality (specific to music), and tone combinatoriality (specific to language).

of figure) and then move on to discuss music (upper part of the figure).

## Language

In thinking about the birth of lexicality in acoustic symbols, I am going to propose that we consider two unconventional though long-established ideas, namely sound symbolism and lexical tone, as well as their union through a "frequency code" in which lexical tones could operate in a sound-symbolic manner (Ohala, 1984, 1994). I find it reasonable to consider the notion of sound symbolism as a potential origin of symbols, a timeworn idea that dates back to the ancient Greeks. Just as gestural theories of language origin are predicated on the idea that gestural pantomimes were the road to achieving gestural symbols (Armstrong and Wilcox, 2007; Arbib, 2012), *so too acoustic pantomimes could have been the road to achieving acoustic symbols* (**Figure 1**). While much research on sound symbolism focuses on phonemic effects related to vowels and consonants (e.g., front vowels connoting small size vs. back vowels connoting large size), a small amount of research relates to what Ohala (1984, 1994) referred to as a "frequency code," in which pitch could be used to iconically convey symbolic meanings. Such a pitch-based code serves as the foundation non-symbolically for affective communication in many animal species and in infant-directed speech, but also has a limited potential to iconically encode spatial features of objects. For example, Nygaard et al. (2009) demonstrated that pitch was effective as a cue to perceive not only emotional valence in speech, but also size and temperature. Perlman et al. (2015) showed that participants could modulate pitch in non-linguistic vocalizations to convey information about vertical position, speed, and texture. Interestingly, sound symbolism has been found to apply to *lexical tone* as well (Ohala, 1984, 1994), with high tones being associated with words conveying small size, and low tones being associated with words conveying large size (so-called size symbolism). While there is no question that arbitrariness ultimately came to dominate the lexicon, it seems reasonable to hypothesize that language evolution got its start by capitalizing on the processing advantages that iconicity could offer (Kita, 2008; Perniss et al., 2010; Imai and Kita, 2011; Perlman and Cain, 2014).

A second hypothesis that I would like to present is that spoken language evolved as a system of lexical tones from its inception (cf Jespersen, 1922), rather than tone being a late emergence. My original model (Brown, 2000a) mistakenly argued that the musilanguage precursor had the property of lexical tone and

thus lexicality, an objection well pointed out by Fitch (2010). I now firmly reject that idea in favor of lexical tone being a purely linguistic feature that emerged *after* the separation of language from the musilanguage precursor. From a cross-linguistic perspective, we know that the majority of spoken languages in the world today are lexical-tonal, although they are concentrated into a handful of geographic hotspots, mainly sub-Saharan Africa, southeast Asia, Papua New Guinea, and the Americas (Yip, 2000). These languages, despite their absence in the well-studied Indo-European language family, represent the dominant mode by which people communicate through speech. Non-tonal languages are the exception, not the rule.

In proposing that language started out as a lexical-tonal system from its origins, I am claiming that the vocal route for developing acoustic symbols involved not just a combinatorial mechanism for phonemes but *a combinatorial mechanism for lexical tones* as well (**Figure 6**). While lexical tone is not conceived of as a combinatorial system by linguists, it seems reasonable to me to think about it this way. Each tone language has a discrete inventory of lexical tones, either level tones (e.g., high, low), contour tones (e.g., rising, falling), or some combination of the two (Yip, 2000). The majority of syllables receive one of these possible tones. Importantly, tone languages contain a large number of homonyms that vary only in tone but in which the phonemes are identical; the four tonal homonyms of /pa/ in Mandarin are a well-cited example of this, where the four words mean eight, to pull out, to hold, and father, respectively (Lee and Zee, 2003). Hence, lexical tone seems to operate similar to the phonemic combinatorial mechanism, but instead works on the pitch levels and/or pitch contours of the vocalic part of the syllable. In other words, while phonemic combinatoriality is principally an articulatory phenomenon, tone combinatoriality is mainly phonatory. An important feature of this hypothesis is that speech inherited and maintained the imprecise levels-and-contours melodic system of the musilanguage system. Lexical tone operates using general pitch contours with imprecise pitch targets (most commonly rising and falling) and likewise with level tones having equally imprecise pitch targets (most commonly high and low). It is critical to keep in mind that, given that lexical tone is absent in one third of contemporary languages, tone is clearly a dispensable feature of a language. However, according to the hypothesis I am offering here, lexical tone is the ancestral state of spoken language, and the loss of tone is a derived feature of non-tonal languages, rather than the reverse progression (Brown, 2000a).

The last feature about the road to language indicated in the lower part of **Figure 6** is the emergence of alternating textures associated with dyadic exchange (see also **Figure 3**). Given that language is about communicating information symbolically, alternation is a much more efficient means of effecting this transmission than simultaneous production, in contrast to music, where simultaneous production is central to its coordinative function and efficacy. I will return to this point about alternation in the section below about the evolution of syntax, since recent work on interactional linguistics demonstrates not only the prosodic relatedness of interacting speakers (Couper-Kuhlen, 2014; Bögels and Torreira, 2015; Filippi, 2016; Levinson, 2016) but their syntactic relatedness as well, leading to models of "dialogic syntax" (Du Bois, 2014; Auer, 2015).

## Music

The road to music is characterized by a complementary set of features emerging from the joint musilanguage precursor. The imprecise nature of the pitch-targets for lexical tone is contrasted with their precision in music and its system of tonality using scaled pitch-levels, which comprises the second major branching from the musilanguage system (**Figure 6**). As paradoxical as it might sound, speech's lexical tones are not the least bit tonal in the musical sense, although both lexical tone and music operate using relative pitch-changes, rather than absolute pitches. Tonality in the musical sense involves a discrete inventory of (relative) pitches making up a musical scale, thereby establishing fixed interval-classes among these pitches, where the same pitches are generally used in both melodic ascent and descent, what I refer to as the *recurrence* of pitches.

Importantly, music is a third type of combinatorial system in human vocal communication (beyond phonemic combinatoriality and lexical-tonal combinatoriality), however in this case involving specific pitch combinations, similar to certain forms of melodious birdsong (Marler, 2000). As mentioned above, while phonemic combinatoriality focuses mainly on articulation, music's pitch combinatoriality focuses mainly on phonation, as with lexical tone. What makes music "musical," and what makes it acoustically different from lexical tone, is that the pitches are scaled and recurrent, whereas in speech, whether for a tone language or an intonation language, they are not. In addition, this scaling of pitch occurs both in the horizontal dimension of melody and in the vertical dimension of harmony (another manifestation of recurrence), since music retains the complex group textures of the precursor stages, although the evolution of rhythmic entrainment mechanisms provides music with a wide diversity of texture types, including human-specific forms of unison and homophony (see **Figure 5** above).

I propose that the coarse-grained levels-and-contours system that was ancestral to speech and music, and that was retained by speech after the bifurcation process, ultimately gave rise to the musical type of tonality, by making a shift from the imprecise pitch-targets of the precursor to the precise intervallic pitch-targets of music. The road to music occurred by a digitization of the pitch properties of the prosodic precursor to produce a scaling of pitches, which serves as the basis for tonality and thus music. For example, as pointed out by the early comparative musicologists (Sachs, 1943), there are simple chants in indigenous cultures that alternate between only two pitches. So, I can imagine scenarios in which the imprecise system of the precursor became quantized so as to settle on recurrent pitch targets through scaling principles.

The final point about the road to music that is indicated in the upper part of **Figure 6** is the emergence of integrated textures associated with group-wide production (see also **Figure 3**). As shown in **Figure 5**, the most integrated textures in human communication are unison and homophony, due to the joint onset of parts. These are the most domain-specific and species-specific textures of music, compared to both human conversation

and animal forms of group vocalizing, where heterophony and polyphony predominate. The emergence of integrated forms of chorusing is due to the advent of mechanisms of not just vocal imitation but metric entrainment (Brown, 2007). While entrainment is often discussed in the literature as the synchronization of movement to some external timekeeper (as in a person tapping their finger to a metronome beat), it occurs comparably as *mutual* entrainment among individuals engaged in chorusing or related forms of synchronized body movement, like marching (Chauvigné et al., 2014).

A major hypothesis of this article is that music evolved to be a *dual* coordination system, using both tonality and metric entrainment to engender integration (**Figure 7**). Scale structure, by digitizing the occurrence of usable pitches, creates *pitch slots* for coordination among chorusing individuals, as manifested in the vertical grid-like pattern of a musical staff, with its discrete pitch levels. Likewise, metrical structure, by creating discrete beat locations for onsets, creates *time slots* for coordination among individuals. The extreme case of integration in music occurs in unison texture—as in the group singing of "Happy Birthday"—where all performers converge on the same pitches at the same time points. However, while music is indeed a dual coordination system, I contend that *scale structure is music's defining feature*, with metrical structure being something that is shared with dance and even with speech in the cases of poetic verse and the rhythmic chanting that occurs at political rallies. In support of this, it is clear that tonality and metrical structure can each work in isolation, as seen both in non-metrical melodies and in metrical structures that are unpitched, such as a tap dance.

An important evolutionary question is how tonality and meter came together to create the dual coordination system that we associate with music. I am inclined to think that entrainment evolved primarily in the context of whole-body synchronization through *dance* (Brown et al., 2006a; Brown and Parsons, 2008; Chauvigné et al., 2014), and that musical chorusing later co-opted this whole-body entrainment system to create musical integration. This jibes with the fact that tonality is domain-specific but that metrical structure is used in a cross-domain fashion, being multi-effector (voice and body) and multi-sensory (we do not need pitch at all for metrical structure or entrainment). Hence, the integrated nature of music emerges from the marriage of a domain-specific pitch system of tonality and a domain-general timing system. Patel (2014) has argued that the trait of metric entrainment evolved jointly with vocal learning, and that the two are casually related to one another. However, I do not agree with that perspective. Since vocal learning is a shared feature between music and speech, I believe it should be placed at the level of the joint vocal precursor described above. In contrast, I see entrainment as emerging outside of this vocal nexus as a system for whole-body coordination through dance, which later gets co-opted by the musical system for use in vocal chorusing and its instrumental analogs (**Figure 7**). To my mind, the relevant co-evolutionary question is not that between entrainment and vocal learning (as per Patel), but instead that between entrainment and *tonality* to create music's dual coordination system.

Before concluding this discussion about the evolution of music, I would like to point out that the evolutionary mystery of music is not just the generation of scale structure *per se*, but the cognitive perception that different scale-types have different emotional valence connotations (Huron, 2006, 2008; Bowling et al., 2012; Parncutt, 2014). I will refer to this as "scale/emotion associations." In Western music theory, there is an association

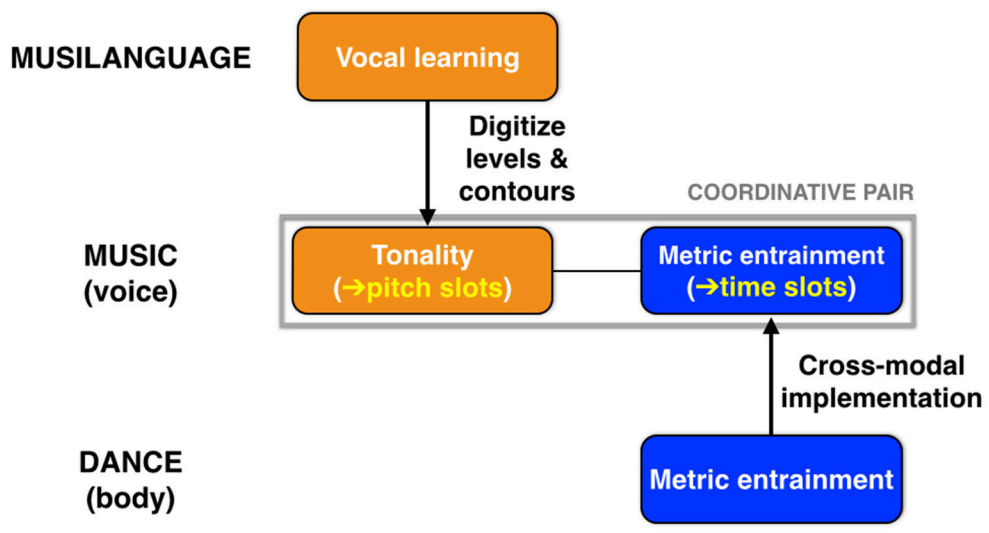



**FIGURE 7 |** Music as a dual coordination system. Music evolved to be a dual coordination system, using both tonality and metric entrainment to engender integration. Scale structure (tonality) provides "pitch slots" for coordination, while metric structure provides "time slots" for coordination. As shown at the top, music inherited vocal learning from the musilangauge precursor, and achieved scale structure through a digitization of the levels-and-contours pitch system of that stage. Metrical structure is proposed to be a cross-modal system that originated as mutual entrainment of whole-body movement during group dancing. Music is able to co-opt this cross-modal system as its rhythmic mechanism.

between the major scale and positive emotional valence, and between the minor scale and negative emotional valence. Scale types can be used in a contrastive manner by composers in a narrative context to convey different emotional meanings, much as contrastive facial expressions can be used by actors to convey different emotions to audiences.

Did scale structure and scale/emotion associations evolve as a unitary phenomenon or instead as two sequential emergences? I could imagine scale structure as serving a coordinative function for group integration all on its own, separate from valence coding, for example as in a musical version of a wolf chorus. Scale/emotion signaling would be more important for emotional expression, by creating a musical language of emotion based on valence coding, involving the contrastive use of two or more scale types. This would be important for group-wide emotional communication. **Figure 8** compares two possible evolutionary models for the emergence of scale structure and scale/emotion processing.

The end result of the musilanguage precursor and its branching to form speech and music is the emergence in humans of a "combinatorial triad" (**Figure 6**) comprised of (1) phonemic combinatoriality for both speech and music, derived from the musilanguage precursor, (2) lexical-tone combinatoriality specific to tone-speech, and (3) pitch combinatoriality specific to music. These systems routinely come together, combining the phonemic and pitch domains. Common examples are the singing of songs with words, such as "Happy Birthday," in which music's tonal properties are combined with speech syllables to musicalize these syllables (discussed in more detail in the last section). But even in the case of singing using vocables (like la-la-la), which is predominant in many world cultures (Lomax, 1968), all singing has to occur using some phoneme or another as the articulatory vehicle, even if it is just a single vowel (as in chanting on /a/) or a nasal consonant (as in humming), although I think that Fitch's (2010) claim that music is bare phonology misses the critical point about tonality.

## SYNTAX EVOLUTION AND THE "PROSODIC SCAFFOLD"

In the introductory section, I argued against a strictly syllabic interpretation of the origin of speech and instead suggested that we need to put emotion, prosody, and communicative intent front and center in our evolutionary thinking, leading me to propose a "prosodic scaffold" perspective. This is the idea that the production of speech is *embedded in prosody*, rather than prosody being an add-on to the compositional and combinatorial levels of speech after the symbolic level of sentence formation has been completed. Importantly, prosody transcends the level of the individual speaker, influencing the process of alternation that characterizes speech's performative arrangement (Robinson, 2005). Recent work on both interactional linguistics and interactional prosody demonstrates the profound influence of this interaction on what people think and say (Couper-Kuhlen, 2014; Du Bois, 2014; Auer, 2015). Speech is not just a process of communication but a process of *coordination*, and prosody serves as both a cause and an effect of this coordination.

**Figure 9** is an expansion of the material shown in **Figure 6** but which now adds the symbolic components of words and sentences. (For ease of interpretation, material from **Figure 6** unrelated to speech is removed). The prosodic scaffold is graphically represented by showing words and sentences (orange color) embedded in prosody (dark red color). At the lowest level of the linguistic hierarchy, the acoustic symbols that comprise individual words are embedded in the context of *word-level prosody*. This would occur through a modulation of the pitch, loudness, duration, and timbral features of constituent syllables to convey both linguistic prosody (e.g., lexical tone, the relative stress of syllables in polysyllabic words) and affective prosody (the valence and intensity of the communicated emotions).

The combination of words to form phrases and sentences brings us to the domain of syntax, without question the most contentious issue in the study of language evolution (Berwick, 2011; Tallerman, 2013). In the previous section, I talked about a "combinatorial triad" for the phonological aspects of speech and music. Syntax too is based on a combinatorial mechanism, but one that is quite different from the ones for phonemes, syllables, lexical tones, and pitches. In this case, it is words that get combined to form sentences, a process of combinatoriality that is referred to as *compositionality* (also productivity). Compared to the small pool of phonemes that go into the combinatorial systems for phonology, the compositional system operates with a pool of tens of thousands of words in the lexicon, organized into word classes that get combined through rule-based syntactic operations to achieve a meaningful ordering of words (Tallerman, 2015).

The field of syntax evolution has witnessed an interesting debate between two contrasting perspectives. The first is the idea of compositionality rooted in the concatenation of symbols, exemplified by "proto-language" models of the type of Bickerton's (1995). The core idea is that, starting from a basic lexicon of individual symbols, these symbols can be combined to form more-complex meanings. At the proto-language stage, the ordering of these symbols is merely associational, and does not suggest any kind of temporal ordering of events or causal relationships, although Jackendoff (1999, 2002) has suggested that Agent First might be a mechanism operative at the proto-language stage. Later stages in the evolution of syntax are thought to add grammaticalization onto the proto-language system to develop word classes that have syntactic functions, not just semantic meanings. Hence, word order and morphology develop, both of which affect the combinability of words as well as their ability to be displaced within sentences. An alternative theory is that speech began from its inception as holistic utterances, called holophrases, and that the evolution of syntax proceeded by fractionating these holophrases into words and phrases (Wray, 1998; Mithen, 2005; Arbib, 2012). The idea is that holophases conveyed complex but holistic meanings, which could be later broken down into constituent words by decomposing their complex meanings. Hence, language evolution proceeded from the holistic level to the unit level, rather than the reverse.

A critical discussion of evolutionary syntax models is beyond the scope of this article. The only point that I will add to the debate is that the "prosodic scaffold" model has the potential to synthesize elements of the two aforementioned classes of theories. The model presented in **Figure 9** integrates combinatorial and holistic processes through the mechanism of prosodic embedding, as shown by "sentence-level prosody" in the figure. Prosody operates in a holistic fashion and is thus inherently holophrastic. The proposal of a prosodic scaffold is that the compositional mechanisms of syntax are *embedded in this holistic prosodic scaffold*. Therefore, instead of arguing for the idea of *symbolic* holophrases, I am arguing for the existence of *prosodic* holophrases that serve as the scaffold for compositional syntactic mechanisms.

Note that this proposal of prosodic embedding is precisely *opposite* to the way that most linguists think about language and its origin. Tallerman (2013:476) states: "Put simply, in syntax words must come first; phrases are built up around single words…Thus, suggesting that phrases evolved in protolanguage *before* there were word classes is once again entirely the wrong way round" (emphasis in the original). This is a difficult argument to address since structural linguists do not consider prosody to be a core component of language. The dispute between linguists and people like me, Mithen and Fitch might have far less to do with our proposals of a co-evolutionary stage uniting language and music as with *how prosodic processes are situated with respect to the core linguistic processes of semantics, syntax, and phonology*. If prosody is linguistically ancillary, then there is no point in discussing a prosodic proto-language that preceded semantics and syntax. If it is primary, then it makes sense to do so. I do not believe that the fields of either linguistics or language evolution have actually had a discussion on this topic.

Leaving aside the idea that language is primarily a vehicle of thought—such as in the monologue that makes up inner speech—language is routinely generated in a discursive manner though the alternating performance arrangement of speech. Sentences must therefore be generated in an interactive manner. But it would be wrong to think of a dialogue as simply a pair of monologues punctuated by interruptions. Sentences are generated during conversation *in response to* what has been said by others (Auer, 2015), not least through the exchange of questions and answers (Jordania, 2006). Language production during conversation, therefore, is a balancing act between two competing needs. On the one hand is the "monologic" driving force of leadership that aims to get one's personal information and perspective across to other people through persuasion, including statements of demands, commands, suggestions, desires, opinions, values, norms, etc. On the other hand is a "dialogic" driving force of mutuality that fosters exchange by *adapting to* one's conversation partner through an ongoing alternation between follower and leader roles. A huge literature on the pragmatics of language (Robinson, 2005) indicates the great extent to which people modify all aspects of linguistic and paralinguistic production so as to adapt to their conversational partners. This occurs at the levels of topic, word choice, syntax, pitch, loudness, tempo, and beyond. The end result of this mutual adaptation is that there is a strong sense of matching, mirroring, and mimicry between conversational partners (Couper-Kuhlen,

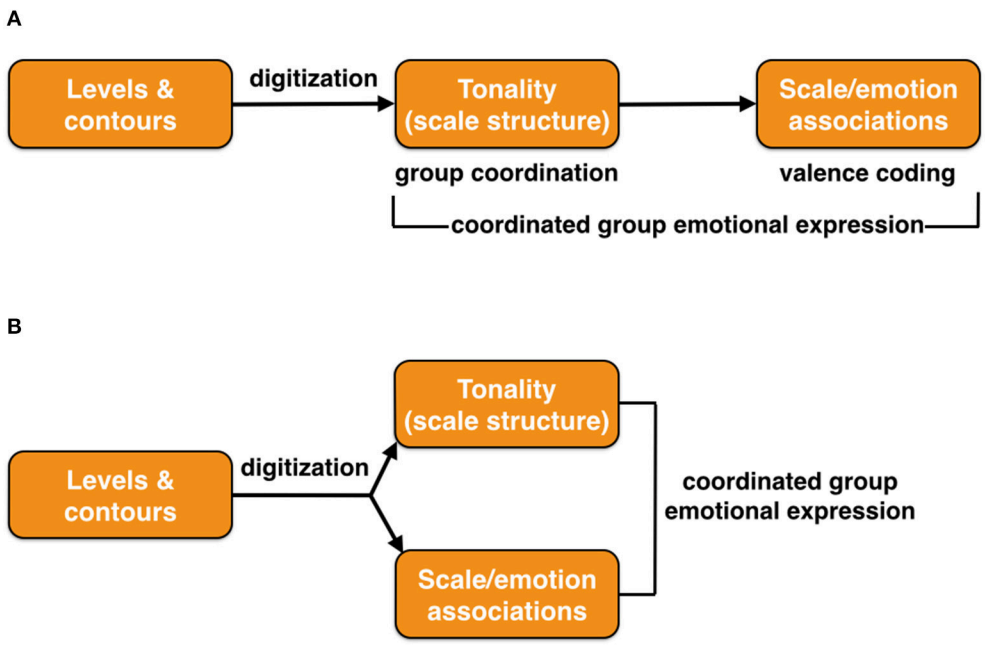



**FIGURE 8 |** The origin of scale/emotion associations in music. Two models are presented for how scale/emotion associations may have evolved, either **(A)** as a sequential occurrence, or **(B)** as a joint occurrence. In both cases, the ultimate outcome is coordinated group emotional expression. The figure is meant to represent an evolutionary timeline, with time progressing from left to right.

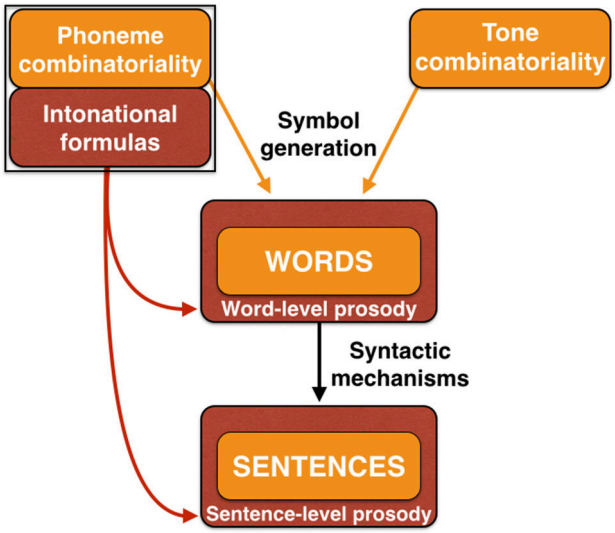



**FIGURE 9 |** A prosodic scaffold in language generation. This figure reproduces the language-relevant material from **Figure 6**, using the same color-coding scheme. Phoneme-combinatorial mechanisms in concert with symbolic-meaning systems generate words as acoustic symbols. In a majority of languages, this is accompanied by the use of lexical tone as well. The "prosodic scafford" model suggests that words are embedded in word-level prosody (dark red color). Sentences are formed through compositional syntactic mechanisms. This process is proposed to be embedded in prosody, in this case sentence-level prosody derived from intonational formulas and affective prosody.

2014), impacting not only language and speech but posture, facial expression, gesture, as well as all aspects of prosody (Szczepek Reed, 2012, 2013; Couper-Kuhlen, 2014; Auer, 2015). Hence, the scaffold of one speaker is clearly influenced by that of another in constructing sentences during conversation.

## (RE)UNIFICATION: THE MUSILINGUISTIC CONTINUUM

A previous section described the bifurcation process to generate language and music as separate, though homologous, functions emanating from a joint prosodic precursor that I called the musilanguage system. As a last step, I now need to consider the (re)unification of language and music (Brown, 2000a), which occurs ubiquitously in the performing arts and religious rituals. The most general interaction between language and music is unquestionably *songs with words*. The potential for direct and seamless coupling between musical pitches and the syllables of words is one of the strongest pieces of evidence for a joint origin of music and language.

However, this coupling does not occur in a singular manner. The comparative musicologist Curt Sachs presciently argued that there was not a unique origin of music but instead multiple (Sachs, 1943). In particular, he proposed a distinction between (1) a type of music derived from melody ("melogenic") and (2) a type of music derived from words and text ("logogenic"). Quite separate from evolutionary considerations *per se*, we can think about Sachs' distinction from a purely structural standpoint and define the melogenic style of singing—whether it occurs with or without words—as being the conventional version of music using scaled intervals and metrically-organized beats. While most forms of melogenic singing in Western culture use words, many others in world cultures do not use words, but instead use vocables (i.e., meaningless syllables like "la" or "heya") as the syllabic vehicles for vocal production.

In contrast to this melogenic style, there are many forms of word-based singing that sound like stylized versions of speech. The focal point of communication is the text, and melody is a means of accentuating it emotively. This logogenic style of singing words is basically a chanting of speech in which the melody and rhythm are closer to speech's intrinsic melody and rhythm than to the melogenic style of scaled pitches and metric rhythms. My interpretation of Sachs' multi-origins hypothesis is that the melogenic style, most notably when using vocables instead of text, arises during the divergence of music from the joint prosodic precursor, and that the logogenic style is something that follows the full-fledged emergence of speech as a cognitive function, where chanting is a means of stylizing the linguistic message.

The argument just described leads me to propose, as I did in Brown (2000a), that the evolutionary processes that generated language and music as reciprocal specializations from a joint precursor ultimately resulted in a "musilinguistic continuum" containing the poles of language and music as well as a number of interactive and intermediate functions (see also Savage et al., 2012). **Figure 10** presents a model of this. At the extremes of the continuum are language and music, represented both vocally (as speech and vocable-based singing, respectively) and instrumentally. The latter includes speech surrogates, such as drummed and whistled languages (Stern, 1957), as well as conventional instrumental music. In the middle of the continuum is the most interactive function of songs with words, where language and music most universally come together (Savage et al., 2015). As shown in the figure, this can be accomplished in a logogenic manner that sounds like a stylized version of speaking, or it can occur in a more melogenic manner, employing musical scales and metric rhythms.

Sitting in between standard speech and songs with words are intermediate functions in which the lexicality of speech is maintained but in which the acoustic properties of the production lean in the direction of music. This can occur with respect to rhythm, melody, or both. Rhythmic speech is a common form of this (Cummins, 2009, 2013), as occurs in poetic verse, rap, and the group chanting that routinely permeates political rallies and marches. Prosodic speech includes the emotionally-accentuated speaking style of an actor, poet, or public speaker, or of a mother interacting with her baby (Fernald, 1989; Papousek, 1996). It also includes logogenic musical forms, such as *sprechstimme*, recitative, and parlando-style chanting, for example cantillation of the Torah. It is important to point out that there are meaningless forms of speech that still have normal speech-like prosodic contours. Examples include the filtered speech used in psychology experiments (Fernald, 1989),

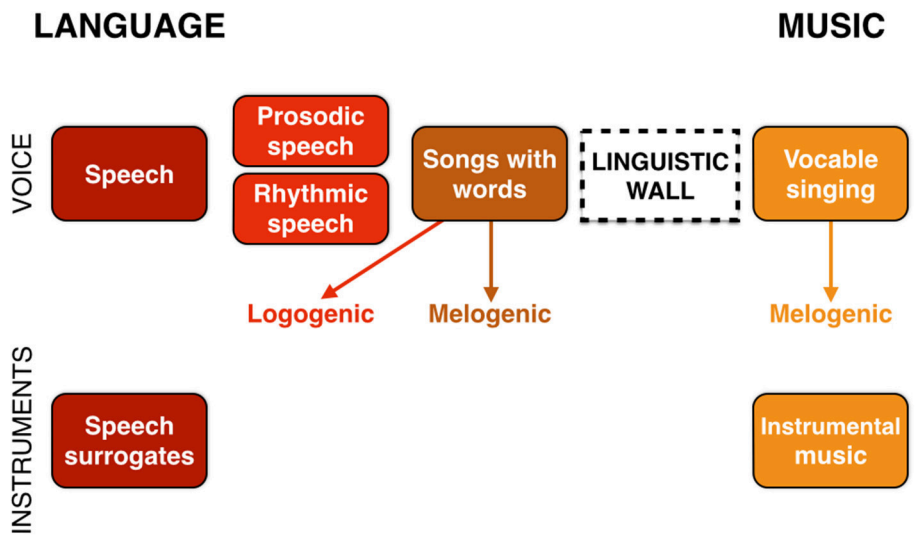



**FIGURE 10 |** The musilingusitic contiuum. At the extremes of the continuum are language and music, represented both vocally and instrumentally. In the middle of the continuum is the most interactive function of songs with words. This can be accomplished in either a logogenic or melogenic manner. Sitting in between standard speech and songs with words are intermediate functions in which the lexicality of speech is maintained but in which the acoustic properties of the production lean in the direction of music, such as rhythmic speech and prosodic speech. There are no comparable intermediate functions on the music side of the continuum since a "linguistic wall" ensures that lexicality is a categorical feature, rather than a continuous acoustic feature like musicality. Vocable singing is shown here as being in the melogenic style. It need not be, but is most commonly found in this form in world musics.

grammelots used in theatrical performance (Jaffe-Berg, 2001; Rudlin, 2015), or simply me watching a film in a foreign language. The point I want to make here is that the elimination of the lexicality of these forms of speech does not suddenly convert them into music. I reject the idea that prosody is a form of music and that prosodic contours divorced of words are a type of singing. Vocalizations like grammelots are better thought of as "de-lexicalized speech" than as music.

While it is easy to talk about intermediate forms of speech that are more or less "musical," I contend that we cannot do the same thing on the musical side of the continuum. As shown by the white box in the figure, I argue that there is a "linguistic wall" that ensures that lexicality is a *categorical* feature, rather than a continuous acoustic feature like musicality. The right side of the figure shows the purely non-lexical functions of vocable singing and instrument playing. If the singing were now to include words instead of vocables, it would immediately jump the linguistic wall to become a song with words. It is difficult to imagine functions in which musicality is maintained but in which lexicality is intermediate between words and non-words. There is really nothing intermediate between words and non-words (e.g., vocables, pseudowords, grunts). So, the linguistic wall creates a categorical divide between vocable singing and word singing, making the musilinguistic continuum both asymmetric in structure and partially discontinuous.

This brings me full circle to the critique that I raised in the opening section of this article of models of language evolution that posit a singing-based precursor stage (Darwin, 1871; Jespersen, 1922). Some people might think that this stage should be identical to what I am calling the musilanguage system in this article. However, I do not see things that way. In particular, the musilanguage stage is proposed to *lack* both tonality and meter, since it is pre-musical. It is a grammelot, hence making it acoustically far more similar to prosodic speech than to music (i.e., it is comprised of levels-and-contours, not scaled intervals). So, while musilanguage might sit next to vocable singing in terms of its absence of lexicality, it would definitely sit next to prosodic speech in terms of its acoustic properties. That is why I find it inappropriate to refer to this precursor as "singing" and why I find it problematic that singing-based theories of language fail to make a distinction between music and prosody. To my way of thinking, tonality (scale structure) is a novel, domain-specific feature of music not shared with speech. That is why I far prefer a neologism like musilanguage to the term singing, since the musical features of singing-based precursors are not specified by people who use the term singing. What they are generally implying is a prosodic vocalization system, rather than a true musical system. I have argued that such a precursor embodies the shared prosodic features of language and music, but not the scales that are specific to music.

## TESTABILITY OF THE MODEL

**Table 1** above lists a dozen proposed features of the musilanguage system. Some of them represent features shared by speech and music, while others do not. For example, much research has shown that affective prosody is conveyed in a parallel manner in speech and music, capitalizing on the same types of dynamic cues (Juslin and Laukka, 2003). By contrast, experimental work from my lab has explored the potential musical properties of speech, and has found that speech is atonal (Chow and Brown, submitted) and based on heterometric rhythms (Brown et al., 2017), both of which conform with properties of the proposed precursor. Likewise, work on singing by Pfordresher and Brown

(2017) suggests that music might in fact be a derivative of a coarse-grained levels-and-contours system, as shown by the highly imprecise nature of sung intervals in everyday singers, not to mention children (Welch, 1979a,b, 2006).

Regarding brain localization, the bifurcation model suggests that music and speech/language should show their greatest similarities at the sensorimotor (phonological) level (Brown et al., 2006b, 2009), but the least similarity with regard to domain-specific features like lexicality and tonality. For example, semantic areas in the inferior and middle temporal gyri are frequently activated during language tasks (Xu et al., 2009; Visser et al., 2012; Krieger-Redwood et al., 2015), but not during music tasks, although these areas can be modulated by music when the task is specifically focused on semantic properties (Koelsch et al., 2004). While the brain network for semantics has been well-studied, that for tonality has been much more poorly explored. A key objective for future research will be to examine the neural basis of what I have called scale/emotion processing, not least the emotional-valence connotations of different scale types. This will unquestionably lead to an exploration of limbic and paralimbic areas associated with emotion perception (Tabei, 2015).

Work on infant development supports the bifurcation model in that the first year of life appears to comprise a shared stage in which prosody, speech, and music are relatively undifferentiated, followed by a separation of language/speech and music/singing as distinct audiovocal functions (Papousek, 1996). Of course, parental singing to/with children in Western culture almost invariably involves the use of songs with words. As a result, most children are taught music through its coupling with language.

The aspect of the model that needs the most verification is the proposal of a "prosodic scaffold" in the production of speech. Work on speech perception demonstrates a strong influence of prosodic cues on comprehending speech (Filippi et al., 2017), but very little work of this type has occurred at the generative level. While numerous studies of affective prosody have used trained actors to convey different basic emotions in speech (Scherer, 2003), no studies have looked at this in spontaneous speech. A mood-induction procedure (Scherer, 2003; Van Dyck et al., 2013) might be one manner to address the influence of affect on speech production, especially if the content of the speech could be controlled for, say through a pre-learned text.

## CONCLUSIONS

The account of language evolution that I have presented in this article is vocal (rather than gestural), prosodic (rather than articulatory or syllabic), group-level (rather than individual, or dyadic), committed to a joint origin of language and music, and rooted in the idea that syntax-based phrase generation emerged, from its origin, as the filling out of a prosodic scaffold during speech production. I propose a two-step evolutionary process: first an involuntary but ritualized system of affective prosody, followed by a learning-based system of intonational prosody grounded in phonemic combinatoriality. From there, language and music branched out as separate, though homologous, functions through the emergence of lexicality and tonality, respectively, and through the adoption of the contrasting communicative arrangements of alternation and integration, respectively. After their separation, language and music are perennially reunited in songs with words, occurring in both melogenic (more-musical) and logogenic (more speech-like) styles. This potential for direct and seamless coupling between words and musical pitches is one of the strongest pieces of evidence supporting a joint origin of language and music.

## AUTHOR CONTRIBUTIONS

SB conceived of the ideas and wrote the manuscript.

## ACKNOWLEDGMENTS

I am grateful to Michel Belyk and Piera Filippi for critical reading of an earlier version of the manuscript and for their useful suggestions for improvement. I thank Aleksey Nikolsky for a detailed and insightful critique of the manuscript, as well as the reviewers for their comments. This work was supported by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

## REFERENCES

Ackermann, H., Hage, S. R., and Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and non-human primates: an evolutionary perspective. *Behav. Brain Sci.* 37, 529–604. doi: 10.1017/S0140525X13003099

Arbib, M. A. (2012). *How the Brain Got Language: The Mirror System Hypothesis*. Oxford: Oxford University Press.

Arboitiz, F. (2012). Gestures, vocalizations, and memory in language origins. *Front. Evol. Neurosci.* 4:2. doi: 10.3389/fnevo.2012.00002

Arboitiz, F., and Garcia, R. (2009). Merging of phonological and gestural circuits in early language evolution. *Rev. Neurosci.* 20, 71–84.

Armstrong, D. F., and Wilcox, S. E. (2007). *The Gestural Origins of Language*. Oxford: Oxford University press.

Auer, P. (2015). "The temporality of language in interaction: Projection and latency," in *Temporality in Interaction*, eds. A. Deppermann and G. Günthner (Amsterdam: John Benjamins Publishing Company), 27–56.

Banse, R., and Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *J. Pers. Soc. Psychol.* 70, 814–836. doi: 10.1037/0022-3514.70.3.614

Beattie, G. (2016). *Rethinking Body Language: How Hand Movements Reveal Hidden Thoughts*. London: Routledge.

Belyk, M., and Brown, S. (2017). The origins of the vocal brain in humans. *Neurosci. Biobehav. Rev.* 77, 177–193. doi: 10.1016/j.neubiorev.2017.03.014

Berwick, R. C. (2011). "Syntax facit saltum redux: biolinguistics and the leap to syntax," in *The Biolinguistic Enterprise*, eds. A. DiSciullo and C. Boeck (Oxford: Oxford University Press), 65–99.

Bickerton, D. (1995). *Language and Human Behavior*. Seattle: University of Washington Press.

Bögels, S., and Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. *J. Phon.* 52, 46–57. doi: 10.1016/j.wocn.2015.04.004

Bowling, D. L., Sundararajan, J., Han, S. E., and Purves, D. (2012). Expression of emotion in Eastern and Western music mirrors vocalization. *PLoS ONE* 7:e31942. doi: 10.1371/journal.pone.0031942

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *J. Zool.* 288, 1–20. doi: 10.1111/j.1469-7998.2012.00920.x

Brown, S. (2000a). "The 'musilanguage' model of music evolution," in *The Origins of Music*, eds N. L. Wallin, B. Merker and S. Brown (Cambridge, MA: MIT Press), 271–300.

Brown, S. (2000b). "Evolutionary models of music: from sexual selection to group selection," in *Perspectives in Ethology. Vol. 13, Behavior, Evolution and Culture*, eds F. Tonneau and N. S. Thompson (New York, NY: Plenum Publishers), 231–281.

Brown, S. (2001). "Are music and language homologues?," in *The Biological Foundations of Music*, eds R. Zatorre and I. Peretz (New York, NY: New York Academy of Sciences), 372–374.

Brown, S. (2007). Contagious heterophony: a new theory about the origins of music. *Music. Sci.* 11, 3–26. doi: 10.1177/102986490701100101

Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltaub, P., and Liotti, M. (2009). The somatotopy of speech: phonation and articulation in the human motor cortex. *Brain Cogn.* 70, 31–41. doi: 10.1016/j.bandc.2008.12.006

Brown, S., Martinez, M. J., and Parsons, L. M. (2006a). The neural basis of human dance. *Cereb. Cortex* 16, 1157–1167. doi: 10.1093/cercor/bhj057

Brown, S., Martinez, M. J., and Parsons, L. M. (2006b). Music and language side by side in the brain. *Eur. J. Neurosci.* 23, 2791–2803. doi: 10.1111/j.1460-9568.2006.04785.x

Brown, S., and Parsons, L. M. (2008). The neuroscience of dance. *Sci. Am.* 299, 32–37. doi: 10.1038/scientificamerican0708-78

Brown, S., Pfordresher, P. Q., and Chow, I. (2017). A musical model of speech rhythm. *Psychomusicology* 27, 95–112. doi: 10.1037/pmu0000175

Chauvigné, L., Gitau, K., and Brown, S. (2014). The neural basis of audiomotor entrainment: an ALE meta-analysis. *Front. Hum. Neurosci.* 8:776. doi: 10.3389/fnhum.2014.00776

Condillac, E. (1746). *An Essay on the Origin of Human Knowledge. English translation by R. G. Weyant (1756), reprinted in facsimile form (1971)*. Gainesville, FL: Scholars' Facsimiles and Reprints.

Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton: Princeton University Press.

Couper-Kuhlen, E. (2014). "Prosody as dialogic interaction," in *Prosody and Phonetics in Interaction*, eds D. Barth-Weingarten and B. Szczepek Reed (Mannheim: Verlag für Gesprächsforschung), 221–251.

Cruttenden, A. (1997). *Intonation, 2nd Edn*. Cambridge: Cambridge University Press.

Cummins, F. (2009). Rhythm as entrainment: the case for synchronous speech. *J. Phon.* 37, 16–28. doi: 10.1016/j.wocn.2008.08.003

Cummins, F. (2013). Joint speech: the missing link between speech and music? *Percepta* 1, 17–32.

Dahlin, C. R., and Benedict, L. (2013). Angry birds need not apply: a perspective on the flexible form and multifunctionality of avian vocal duets. *Ethology* 120, 1–10. doi: 10.1111/eth.12182

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. London: J. Murray.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. New York, NY: Oxford University Press.

Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *J. Acoust. Soc. Am.* 129, 2245–2252. doi: 10.1121/1.3562174

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013

Dingemanse, M., Schuerman, W., Reinisch, E., Rufvesson, S., and Mitterer, H. (2016). What sound symbolism can and cannot do: testing the iconicity of ideophones from five languages. *Language* 92, e117–e133. doi: 10.1353/lan.2016.0034

Dissanayake, E. (2000). "Antecedents of the temporal arts in early mother-infant interaction," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 389–410.

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press.

Du Bois, J. W. (2014). Towards a dialogic syntax. *Cogn. Linguist.* 25, 359–410. doi: 10.1515/cog-2014-0024

Falk, D. (2009). *Finding Our Tongues: Mothers, Infants, and the Origins of Language*. New York, NY: Basic Books.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Dev.* 60, 1497–1510. doi: 10.2307/1130938

Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Front. Psychol.* 7:1393. doi: 10.3389/fpsyg.2016.01393

Filippi, P., Ocklenburg, S., Bowling, D. L., Heege, L., Güntürkün, O., Newen, A., et al. (2017). More than words (and faces): evidence for a Stroop effect of prosody in emotion word processing. *Cogn. Emot.* 31, 879–891. doi: 10.1080/02699931.2016.1177489

Fitch, W. T. (2010). *Evolution of Language*. Cambridge: Cambridge University Press.

Fuchs, S., Petrone, C., Krivokapic, J., and Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *J. Phon.* 41, 29–47. doi: 10.1016/j.wocn.2012.08.007

Garcia, R. R., Zamorao, F., and Aboitiz, F. (2014). From imitation to meaning: circuit plasticity and the acquisition of a conventionalized semantics. Front. Hum. *Neurosci.* 8:605. doi: 10.3389/fnhum.2014.00605

Ghazanfar, A. A., Takahashi, D. Y., Mathur, N., and Fitch, W. T. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Curr. Biol.* 22, 1176–1182. doi: 10.1016/j.cub.2012.04.055

Grosjean, F., and Collins, M. (1979). Breathing, pausing and reading. *Phonetica* 36, 98–114. doi: 10.1159/000259950

Hagen, E. H., and Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Hum. Nat.* 14, 21–51. doi: 10.1007/s12110-003-1015-z

Halliday, M. A. K. (1967). *Intonation and Grammar in British English*. The Hague: Mouton.

Halliday, M. A. K. (1970). *A Course in Spoken English: Intonation*. London: Oxford University Press.

Hockett, C. (1960). The origin of speech. *Sci. Am.* 203. 88–111. doi: 10.1038/scientificamerican0960-88

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Huron, D. (2008). A comparison of average pitch height and interval size in major- and minor-key themes: evidence consistent with affect-related pitch prosody. *Empir. Musicol. Rev.* 3, 59–63. doi: 10.18061/1811/31940

Imai, M., and Kita, S. (2011). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130298. doi: 10.1098/rstb.2013.0298

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends Cogn. Sci.* 3, 272–279. doi: 10.1016/S1364-6613(99)01333-9

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackendoff, R. (2011). What is the human language faculty? Two views. *Language* 87, 586–624. doi: 10.1353/lan.2011.0063

Jaffe-Berg, E. (2001). Forays into grammelot: the language of nonsense. *J. Dramatic Theory Crit.* 15, 3–15.

Jespersen, O. (1922). *Language: Its Nature, Development and Origin*. New York, NY: Norton.

Jordania, J. (2006). *Who Asked the First Question? The Origins of Human Choral Singing, Intelligence, Language and Speech*. Tblisi: Logos.

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Keating, P., and Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Work. Pap. Phon.* 101, 112–156.

Kita, S. (2008). "World-view of protolanguage speakers as inferred from semantics of sound symbolic words: a case of Japanese mimetics," in *The Origins of Language: Unraveling Evolutionary Forces*, ed M. Masataka (Cham, CH: Springer), 25–38.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nat. Neurosci.* 7, 302–307. doi: 10.1038/nn1197

Krieger-Redwood, K., Teige, C., Davey, J., Hymers, M., and Jefferies, E. (2015). Conceptual control across modalities: graded specialisation for pictures and words in inferior frontal and posterior temporal cortex. *Neuropsychologia* 76, 92–107. doi: 10.1016/j.neuropsychologia.2015.02.030

Krivokapic, J. (2007). Prosodic planning: effects of phrasal length and complexity on pause duration. *J. Phon.* 35, 162–179. doi: 10.1016/j.wocn.2006.04.001

Krivokapic, J. (2012). "Prosodic planning in speech production," in *Speech Planning and Dynamics*, eds S. Fuchs, M. Weihrich, D. Pape, and P. Perrier (Bern, CH: Peter Lang), 157–190.

Ladd, D. R. (2008). *Intonational Phonology, 2nd Edn*. Cambridge: Cambridge University Press.

Ladd, D. R. (2012). What is duality of patterning, anyway? *Lang. Cogn.* 4, 261–273. doi: 10.1515/langcog-2012-0015

Lee, W. S., and Zee, E. (2003). Standard Chinese (Beijing). *J. Int. Phon. Assoc.* 33, 109–112. doi: 10.1017/S0025100303001208

Lerdahl, F. (2001). The sounds of poetry viewed as music. *Ann. N. Y. Acad. Sci.* 930, 337–354. doi: 10.1111/j.1749-6632.2001.tb05743.x

Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Levelt, W. J. (1999). Models of word production. *Trends Cogn. Sci.* 3, 223–232. doi: 10.1016/S1364-6613(99)01319-4

Levinson, S. C. (2016). Turn-taking in human communication: origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Lomax, A. (1968). *Folk Song Style and Culture.* New Brunswick: American Association for the Advancement of Science.

MacLarnon, A., and Hewitt, G. (1999). The evolution of human speech. *Am. J. Phys. Anthropol.* 109, 341–363. doi: 10.1002/(SICI)1096-8644(199907)109:3<341::AID-AJPA5>3.0.CO;2-2

MacLarnon, A., and Hewitt, G. (2004). Increased breathing control: another factor in the evolution of human language. *Evol. Anthropol.* 13, 181–197. doi: 10.1002/evan.20032

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Beh. Brain Sci.* 21, 499–546. doi: 10.1017/S0140525X98001265

MacNeilage, P. F., and Davis, B. L. (2005). The frame/content theory of evolution of speech: a comparison with a gestural-origins alternative. *Interact. Stud.* 6, 73–99. doi: 10.1075/is.6.2.03mac

Malm, W. P. (1996). *Music Cultures of the Pacific, the Near East, and Asia, 3rd Edn*. Upper Saddle River, NJ: Prentice Hall.

Margulis, E. H. (2013). Repetition and emotive communication in music versus speech. *Front. Psychol.* 4:167. doi: 10.3389/fpsyg.2013.00167

Margulis, E. H. (2014). *On Repeat: How Music Plays the Mind*. New York, NY: Oxford University Press.

Marler, P. (2000). "Origins of music and speech: insights from animals," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 31–48.

McGinn, C. (2015). *Prehension: The Hand and the Emergence of Humanity*. Cambridge, MA: MIT Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

Mertens, P. (2004). "The Prosogram: semi-automatic transcription of prosody based on a tonal perception model," in *Proceedings of the Speech Prosody conference* 2004 (Nara).

Mithen, S. J. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld and Nicolson.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., and Kirby, S. (2016). How arbitrary is language? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130299. doi: 10.1098/rstb.2013.0299

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* 111, 855–869. doi: 10.1086/283219

Nadel, E. (1930). The origins of music. *Musical Q.* 16, 531–546. doi: 10.1093/mq/XVI.4.531

Newman, E. (1905). *Musical Studies*. New York, NY: Haskell House Publishers.

Nikolsky, A. (2015). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.01405

Nikolsky, A. (2016). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-2: ancient to seventeenth century. *Front. Psychol.* 7:211. doi: 10.3389/fpsyg.2016.00211

Nygaard, L. C., Herold, D. S., and Lamy, L. L. (2009). The semantics of prosody: acoustic and perceptual evidence of prosodic correlates to word meaning. *Cogn. Sci.* 33, 127–146. doi: 10.1111/j.1551-6709.2008.01007.x

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica* 41, 1–16. doi: 10.1159/000261706

Ohala, J. J. (1994). "The frequency code underlies the sound-symbolic use of vocal pitch," in *Sound Symbolism*, eds J. Hinton, J. Nichols and J. J. Ohala (Cambridge: Cambridge University Press), 325–247.

Papousek, M. (1996). "Intuitive parenting: a hidden source of musical stimulation in infancy," in *Musical Beginnings: Origins and Development of Musical Competence*, eds Deliège and J. Sloboda (Oxford: Oxford University Press), 88–112.

Parncutt, R. (2014). The emotional connotations of major versus minor tonality: one or more origins? *Mus. Sci.* 18, 324–353. doi: 10.1177/1029864914542842

Patel, A. D. (2008). *Music, Language and the Brain*. Oxford: Oxford University Press.

Patel, A. D. (2014). The evolutionary biology of musical rhythm: was Darwin wrong? *PLoS Biol.* 12:e1001821. doi: 10.1371/journal.pbio.1001821

Perlman, M., and Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture* 14, 320–350. doi: 10.1075/gest.14.3.03per

Perlman, M., Dale, R., and Lupyan, G. (2015). Iconicity can ground the creation of vocal symbols. *R. Soc. Open Sci.* 2:150152. doi: 10.1098/rsos.150152

Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* 1:227. doi: 10.3389/fpsyg.2010.00227

Pfordresher, P. Q., and Brown, S. (2017). Vocal mistuning reveals the nature of musical scales. *J. Cogn. Psychol.* 29, 35–52. doi: 10.1080/20445911.2015.1132024

Pickett, J. M. (1999). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Needham Heights, MA: Allyn and Bacon.

Provine, R. R. (2017). Laughter as an approach to vocal evolution: the bipedal theory. *Psychon. Bull. Rev.* 24, 238–244. doi: 10.3758/s13423-016-1089-3

Robinson, D. (2005). *Introducing Performative Pragmatics*. New York, Ny: Routledge.

Rousseau, J.-J. (1781). *Essay on the Origin of Languages. English translation by J. H. Moran and A. Gode (1986)*. Chicago, IL: University of Chicago Press.

Rudlin, J. (2015). "Grommelot," in *The Routledge Companion to Commedia dell'Arte*, eds J. Chaffee and O. Crick (New York, NY: Routledge), 155–164.

Sachs, C. (1943). *The Rise of Music in the Ancient World: East and West*. London: J. M. Dent and Sons.

Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U.S.A* 112, 8987–8992. doi: 10.1073/pnas.1414495112

Savage, P. E., Merritt, E., Rzeszutek, T., and Brown, S. (2012). CantoCore: a new cross-cultural song classification scheme. *Anal. Approaches World Music* 2, 87–137.

Scherer, K. R. (2003). Vocal communication of emotion: a review of research paradigms. *Speech Comm.* 40, 227–256. doi: 10.1016/S0167-6393(02)00084-5

Scott-Phillips, T. (2017). Pragmatics and the aims of language evolution. *Psychon. Bull. Rev.* 24, 186–189. doi: 10.3758/s13423-016-1061-2

Slater, P. (2000). "Birdsong repertoires: their origins and use," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 49–63.

Spencer, H. (1857). The origin and function of music. *Frasers Mag.* 56, 396–408.

Spencer, H. (1890). The origin of music. *Mind* 15, 449–468. doi: 10.1093/mind/os-XV.60.449

Stern, T. (1957). Drum and whistle "languages": an analysis of speech surrogates. *Am. Anthropol.* 59, 487–506. doi: 10.1525/aa.1957.59.3.02a00070

Svantesson, J.-O. (2017). Sound symbolism: the role of word sound in meaning. *WIREs Cogn. Sci.* 8:e1441. doi: 10.1002/wcs.1441

Szczepek Reed, B. B. (2012). Beyond the particular: prosody and the coordination of actions. *Lang. Speech* 55, 13–34. doi: 10.1177/0023830911428871

Szczepek Reed, B. B. (2013). "Suprasegmentals: prosody in conversation," in *The Encyclopedia of Applied Linguistics*, ed C. A. Chapelle (Oxford: Wiley-Blackwell), 1–6.

Tabei, K.-I. (2015). Inferior frontal gyrus activation underlies the perception of emotions, while precuneus activation underlies the feeling of emotions during music listening. *Behav. Neurol.* 2015:529043. doi: 10.1155/2015/529043

Tallerman, M. (2013). Join the dots: a musical interlude in the evolution of language? *J. Linguist.* 49, 455–487. doi: 10.1017/S0022226713000017

Tallerman, M. (2015). *Understanding Syntax, 4th Edn*. London: Routledge.

Thomas, D. A. (1995). *Music and the Origins of Language: Theories from the French Enlightenment*. Cambridge: Cambridge University Press.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Van Dyck, E., Maes, P.-J., Hargreaves, J., Lesaffre, M., and Leman, M. (2013). Expressing induced emotions through free dance movement. *J. Nonverbal Behav.* 37, 175–190. doi: 10.1007/s10919-013-0153-1

Visser, M., Jefferies, E., Embleton, K. V., and Lambon Ralph, M. A. (2012). Both the middle temporal gyrus and the ventral anterior temporal area are crucial for multimodal semantic processing: distortion-corrected fMRI evidence for a double gradient of information convergence in the temporal lobes. *J. Cogn. Neurosci.* 24, 1766–1778. doi: 10.1162/jocn_a_00244

Wallaschek, R. (1891). On the origins of music. *Mind* 16, 375–386. doi: 10.1093/mind/os-XVI.63.375

Welch, G. (2006). "Singing and vocal development," in *The Child as Musician: A Handbook of Musical Development*, ed G. McPherson (Oxford: Oxford University Press), 311–329.

Welch, G. F. (1979a). Poor pitch singing: a review of the literature. *Psychol. Music* 7, 50–58. doi: 10.1177/030573567971006

Welch, G. F. (1979b). Vocal range and poor pitch singing. *Psychol. Music* 7, 13–31. doi: 10.1177/030573567972002

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Lang. Commun.* 18, 47–67. doi: 10.1016/S0271-5309(97)00033-5

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., and Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20664–20669. doi: 10.1073/pnas.0909197106

Yip, M. (2000). *Tone*. Cambridge: Cambridge University Press.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Brown. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 06 November 2017
doi: 10.3389/fpsyg.2017.01820

# The Paradox of Isochrony in the Evolution of Human Rhythm

*Andrea Ravignani*[1,2,3]* *and Guy Madison*[4]

[1] *Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands,* [2] *Veterinary and Research Department, Sealcentre Pieterburen, Pieterburen, Netherlands,* [3] *Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium,* [4] *Department of Psychology, Umeå University, Umeå, Sweden*

Isochrony is crucial to the rhythm of human music. Some neural, behavioral and anatomical traits underlying rhythm perception and production are shared with a broad range of species. These may either have a common evolutionary origin, or have evolved into similar traits under different evolutionary pressures. Other traits underlying rhythm are rare across species, only found in humans and few other animals. Isochrony, or stable periodicity, is common to most human music, but isochronous behaviors are also found in many species. It appears paradoxical that humans are particularly good at producing and perceiving isochronous patterns, although this ability does not conceivably confer any evolutionary advantage to modern humans. This article will attempt to solve this conundrum. To this end, we define the concept of isochrony from the present functional perspective of physiology, cognitive neuroscience, signal processing, and interactive behavior, and review available evidence on isochrony in the signals of humans and other animals. We then attempt to resolve the paradox of isochrony by expanding an evolutionary hypothesis about the function that isochronous behavior may have had in early hominids. Finally, we propose avenues for empirical research to examine this hypothesis and to understand the evolutionary origin of isochrony in general.

**Keywords: synchrony, prediction, interaction, coordination, turn-taking, evolution of music, evolution of speech**

OPEN ACCESS

***Edited by:***
*Leonid Perlovsky,*
*Harvard University and Air Force Research Laboratory, United States*

***Reviewed by:***
*Shinya Fujii,*
*Keio University, Japan*
*Enrico Glerean,*
*Aalto University, Finland*

***Correspondence:**
*Andrea Ravignani*
*andrea.ravignani@gmail.com*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology*

***Received:*** *25 May 2017*
***Accepted:*** *30 September 2017*
***Published:*** *06 November 2017*

***Citation:***
*Ravignani A and Madison G (2017) The Paradox of Isochrony in the Evolution of Human Rhythm. Front. Psychol. 8:1820. doi: 10.3389/fpsyg.2017.01820*

## CONTENTS

This paper deals with isochronous temporal patterns. The emphasis is on the quantitative properties of isochronous patterns, and their perception and production in humans. The paper is organized in seven sections, namely:

(1) *What is isochrony?*, where we lay out crucial definitions and summarize basic relevant concepts;
(2) *The relevance of isochrony to human music and speech*, where we discuss how isochrony might partly underlie some behaviors in modern humans, such as music, speech and dance;
(3) *Mathematics, physics and signal processing*, where we discuss isochrony from the pure perspective of its physical and mathematical structure (as opposed, for instance, to its biological, behavioral or cognitive nature);
(4) *Physiology and neuroscience*, where we suggest how isochronous patterns have biological and psychological relevance for living organisms;
(5) *Comparative cognition: Non-human animals*, where we briefly summarize previous empirical attempts in finding, either directly or indirectly, isochronous behaviors in other species;

(6) *Isochrony in interaction*, where we move from isochronous behaviors in single individuals to group behaviors potentially involving isochrony;
(7) *Evolutionary hypotheses and future empirical work*, where we join all strands laid out in the previous six sections, and sketch an evolutionary account for the origin of isochrony in our species.

The aim of this paper is not to provide an exhaustive review of each of these areas. Rather, we attempt to establish a first connection between as many explanatory levels of isochrony as possible, across scientific disciplines and research traditions.

## WHAT IS ISOCHRONY?

Music is a complex phenomenon composed of interdependent parts. While a holistic approach is always important, the analytic, scientific method works by first analyzing constituent components individually (Fitch, 2015). Rhythm is a crucial dimension of human music. In common language, but also in scientific publications, different meanings are often conflated into the word 'rhythm' (see **Table 1** for definitions). In the most general definition, *rhythm* denotes a pattern of events in time (McAuley, 2010). An *isochronous* pattern is a rhythm where all intervals between events are equal, like those of a metronome. Hence, all isochronous sequences are rhythmic, but not vice-versa (**Figure 1**). A third related concept is the *beat*, namely the psychological tendency to superimpose an isochronous grid to a rhythmic sequence (a.k.a. pulse or beat perception, **Figure 2C**). The focus of the present paper is the evolutionary significance of the human perception and production of rhythmic sequences that are physically isochronous, henceforth simply 'isochronous.' We deliberately avoid discussing pulse and beat perception, as these have been object of much empirical research and many theoretical frameworks. Pure isochrony has received comparatively less attention.

Humans are particularly good at producing and perceiving both rhythmic and isochronous patterns (Bolton, 1894; Fitch, 2009, 2015; Motz et al., 2013; Ravignani et al., 2017). Yet, while general rhythm capacities could be biologically useful for foraging, mating, navigating the environment and predicting events, propensity for isochrony does not seem to confer any evolutionary advantage to modern humans (Fitch, 2009). Humans' aptitude for isochrony contrasted with its apparent lack of evolutionary function constitutes the paradox of isochrony. Below we offer some perspectives on the presence of isochrony in nature and in modern humans' everyday life.

Humans have extraordinary abilities to deal with isochronous behaviors. We can detect deviations from isochrony on the order of 20 ms or 4% of the interval to be timed, but we can also perceive the underlying isochrony even in the face of random deviations and gradual increases/decreases in intervals (Bolton, 1894; Madison and Merker, 2002, 2004; Max and Yudman, 2003; Repp, 2005; Madison, 2009; Motz et al., 2013; Repp and Su, 2013). When confronted with isochronous sequences where intervals have been slightly jittered, humans will tend

**TABLE 1 |** Definitions of key concepts discussed in the paper (Expanded and modified from Ravignani and Norton, 2017).

| Name | Definition |
|---|---|
| Rhythm | Pattern of events in time; i.e., 'a specific succession of durations' (Wade, 2004, p. 57). |
| Interval | Temporal duration encompassed by two events. |
| Idealized isochrony | A rhythmic pattern where all intervals have *equal duration*. |
| Isochrony (empirical) | A rhythmic pattern where all intervals have *roughly* equal duration. Alternatively, a rhythmic pattern obtained by jittering events in an idealized isochronous pattern. |
| Anisocronous | Not isochronous. Operationally, a pattern exhibiting less isochrony than another pattern under consideration. |
| Pulse a.k.a. beat | Psychological tendency to superimpose an isochronous grid (not necessarily present in the physical signal) to a rhythmic sequence. In some cases, such as here, pulse and beat are defined as equivalent. In some other cases, the beat is defined as the isochronous grid generated via metrical expectations (see below). |
| Synchronous | Two rhythmic patterns where each event in one pattern occurs at the same time as a corresponding event in the other pattern. |
| Meter | Hierarchical organization of temporal events based on stress and other spectral properties, such as loudness alternation, pitch variation, etc. |
| Syllable-timed | Language where all accented and unaccented syllables are isochronous. |
| Stress-timed | Language where all stressed syllables occur isochronously. |
| Mora-timed | Language where all moras, consisting in syllables or some syllable combinations, are isochronous. |
| Endogenous | Pattern production involving a 'self-sustained clock that governs the rhythm' (Pikovsky et al., 2003, p. 87). |
| Exogenous | Pattern production that is influenced by an *external* oscillator/clock/pacemaker. |
| Reactive | Exogenous mechanism of production of a temporal interval whose duration is a direct *response* from a previous duration in the environment. |
| Predictive | Exogenous mechanism of production of a temporal interval whose duration is based on *expectations* about an upcoming environmental duration. |
| Entropy | The expected value of the information content of a signaling system. It quantifies the *amount of information* in a signal, without any reference to its meaning. |
| Redundancy | A measure of the actual information content relative to the maximum possible information content of a signaling system. High redundancy means that information is more 'diluted.' |
| Compressibility | Here: inverse of entropy. |
| Expressivity | Capacity of a signaling system to convey different *meanings*. |

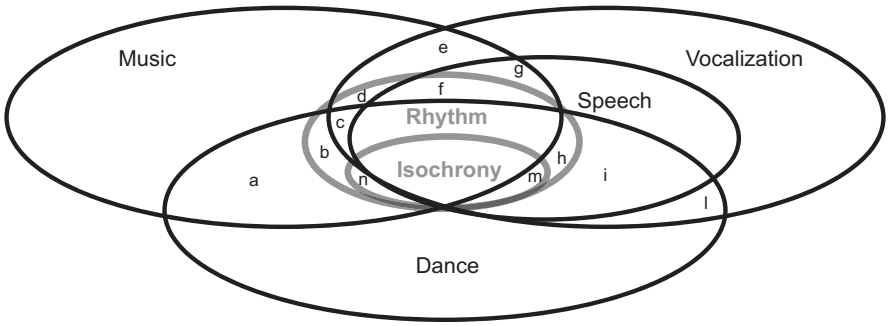



**FIGURE 1 |** Conceptual relationships among music, vocalizations, speech, and dance with respect to temporal structure, rhythm, and isochrony. Lowercase letters a–n denote some possible intersections. For instance, *g* denotes shared timing between speech and music (but not dance), while *n* denotes isochronous music and dance with no commonalities with speech. It is up to empirical research to decide which ones of these are empty. Notice that the question of intersections is not moot: one of the first papers on the origins of dance (Laland et al., 2016), for instance, implicitly suggested that the evolutionary path for the emergence of dance had to go from *e* to *m* to *h*.

to regularize the intervals and perceive the whole sequence as isochronous (for psychophysical thresholds see Friberg and Sundberg, 1995; Madison and Merker, 2002; Merker et al., 2009). Also when producing isochronous sequences the variability is around 4%, but it can be decreased to about 2% of the interval when synchronizing to a sound sequence with faster metrical levels (Madison, 2014). Listening to an isochronous sequence typically induces a beat (**Figure 2C**), which in turn generates expectations that future events will fall into a multiple or integer subdivision of the beat. A violation of the expectation, such as slightly changing the onset of one event, can be measured by mismatch negativity (Motz et al., 2013). The processes underlying this reaction are subliminal (Madison and Merker, 2004) and require no learning, which indicates that the beat is a very basic, inherited phenomenon.

In evolutionary biology, a behavioral trait can appear for a number of reasons. It can be an evolutionary *adaptation*, namely a trait which evolved to increase a species' fitness in a given environment. As such, it may more or less have lost its adaptive value due to changes in the environment, while still prevailing in the population because it has not been selected against. It can also be a by-product of other evolutionary processes, a so-called *exaptation*. As such, isochrony might have been recruited for purposes unrelated to the pressures which caused its early emergence (Merker et al., 2009). In the present context, isochrony refers to humans' perception and production of isochronous event sequences within the bounds and constraints reviewed above.

## THE RELEVANCE OF ISOCHRONY TO HUMAN MUSIC AND SPEECH

The roots of the human propensity for isochrony are clearly found in our biology, specifically in some preparedness of our neural system (Buzsaki, 2006; Arnal and Giraud, 2012; Fujioka et al., 2012; Fujii and Wan, 2014; Merchant et al., 2015). For example, newborn babies react differently to isochronous than to anisochronous sequences (Honing et al., 2009). Similarly, children aged 2–4 years show motoric isochronous behavior with clear periodicities, though little tempo adjustment (Eerola et al., 2006). Although isochrony in music is a human universal (**Figure 2C**), there is considerable variation across the worlds' music cultures. Western musical cultures appear to employ isochrony most thoroughgoingly, for example with rhythmic sequences being composed of isochronous subsequences (e.g., Bach). When rhythmic patterns are not isochronous, they are based on a psychologically induced sense of beat (Merker et al., 2009). In these cases, notes that are played continuously confirm or violate the induced pulse (**Figure 2D**), either in a structural or expressive manner (see Merker, 2014 for a novel perspective). African music is also isochronous at some descriptive level (though see **Figure 3**), while Asian music tends to be less so. Some North-American Indian, Javanese Gamelan, and Western electro-acoustic traditions exhibit no isochrony at all, but it might be argued that they do not fulfill reasonable definitions of music. For comparisons of timing in different musical cultures, see (Arom, 1991; Polak et al., 2016; Neuhoff et al., 2017). Although humans are cognitively biased toward isochrony in music (Ravignani et al., 2016a; Fitch, 2017), this bias is apparently modulated by enculturation (Jacoby and McDermott, 2017, though see Bowling et al., 2017). Finally, isochrony is often associated with motor synchronization in the literature. However, a recent medical case study has found a dissociation between perception of isochrony, among others, and audio-motor synchronization abilities (Bégel et al., 2017).

Speech is another human activity which may involve isochrony (**Figure 1**). The research field investigating rhythmic regularities in speech has been split for decades (Lehiste, 1977; Roach, 1982; Kotz and Schwartze, 2010, 2016; Fujii and Wan, 2014). Some scholars argue that world languages can be classified in groups exhibiting isochrony at the sentence, mora, or syllable levels (called, respectively, stress-timed, mora-timed, and syllable-timed languages, see **Table 1** and Grabe and Low, 2002; Fabb and Halle, 2012). Other researchers argue the opposite,

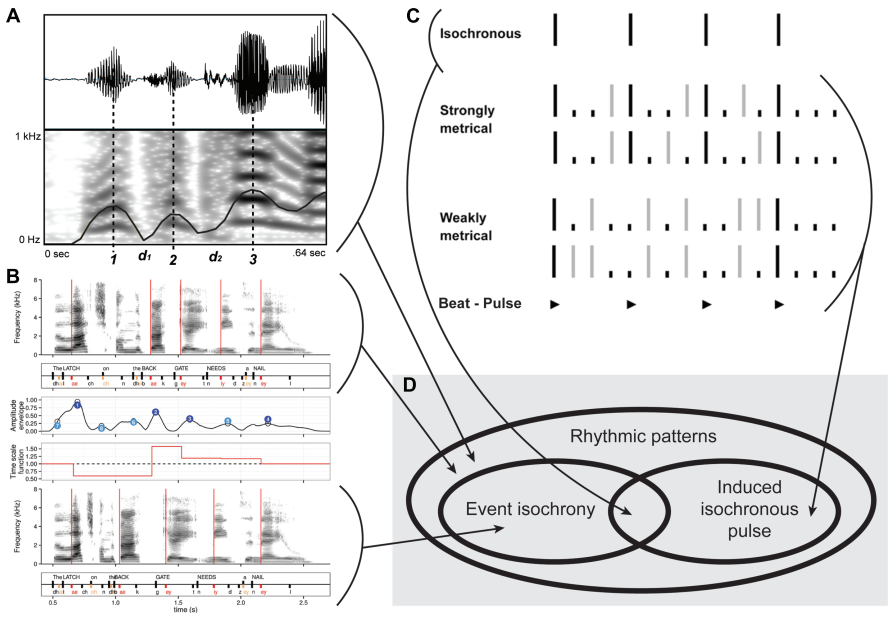



**FIGURE 2 |** Rhythm, isochrony and pulse. **(A)** Examples of a non-isochronous speech rhythmic pattern. Figure modified from Jadoul et al. (2016). **(B)** Examples of another non-isochronous speech rhythmic pattern **(top)**, manipulated into isochrony **(bottom)**. Figure modified from Aubanel et al. (2016). **(C)** Musical rhythm showing event isochrony **(top)** and non-isochronous musical pattern generating isochronous expectations **(bottom)**. Figure modified from Launay et al. (2014). **(D)** Venn diagram showing the relationship among: rhythmic patterns vs. event isochrony vs. isochronous pulse perception.

namely that the speech signal is inherently anisochronous (**Figures 2A,B**), and the feeling of isochrony derives, for instance, from perceptual regularization rather than physical properties of the signal (Tuller and Fowler, 1980; Dauer, 1983; Jadoul et al., 2016; Brown et al., 2017). Without entering this debate here, some empirical findings are worth noticing. In particular, no matter the theoretical perspective adopted, human *vocalizations* can be experimentally driven toward isochrony (Jacoby and McDermott, 2017), especially when two individuals are asked to speak synchronously (Bowling et al., 2013) or perform turn-taking (Schultz et al., 2016). If speech recordings are experimentally manipulated, so that the syllable timing follows heterogeneous rhythmic patterns, isochronous speech is more intelligible than anisochronous speech (Aubanel et al., 2016). Finally, a recent experiment found evidence for isochronous timing in children's handwriting (Pagliarini et al., 2017).

A third common human activity where isochrony is hypothesized to play a role is dance (Fitch, 2016; Laland et al., 2016; Richter and Ostovar, 2016; Su, 2016a,b). Similarly to music, a series of isochronous events, such as a drum line, may provide anchor points in time used to structure dance movements (Fitch, 2016; Laland et al., 2016). Likewise, biophysical constraints on movement produce isochronous or integer ratio temporal intervals (e.g., Merker et al., 2009; Su, 2016a). This isochrony-centered perspective might however be quite specific for dance in humans inhabiting the Western world. A more inclusive approach considers dance a polyhedric behavior present in all human cultures, and whose precursors can be found in other animal species (Fink and Shackelford, 2016; Ravignani and Cook, 2016). If this approach is adopted, then isochrony might not be such an indispensable pillar of dance (Ravignani and Cook, 2016).

All the above suggests that, while isochrony might not be crucial in dance or speech, it is present in human everyday musical behavior (**Figure 1**). So, why is isochrony so common if it doesn't appear to serve any particular function (Fitch, 2009; Merker et al., 2009), at least in modern humans? Below we will try to analytically decompose isochrony even further in its constituent parts across disciplines.

## MATHEMATICS, PHYSICS, AND SIGNAL PROCESSING

Rhythms, including isochronous ones, can be formalized mathematically (Cohen, 1962; Toussaint, 2013). From a purely information-theoretic perspective (MacKay, 2003), when producing a signal over time, isochrony minimizes the signal's entropy. A pattern of time intervals (**Figure 4A**) can be described by a set of interval durations and the probability of occurrence of each interval (**Figure 4B**). A more refined model features conditional transition probabilities (**Figures 4C,D**), where the duration of the upcoming interval is determined probabilistically from the duration of the previous *n* intervals (Cohen, 1962). An example could be a specific pattern for which, given that the first and second intervals are short, there is a high probability that the third interval is long. This is not the case for isochronous sequences: no matter which interval is to be predicted, and how many past intervals are taken into account, the upcoming interval will be a constant value equal to past intervals with

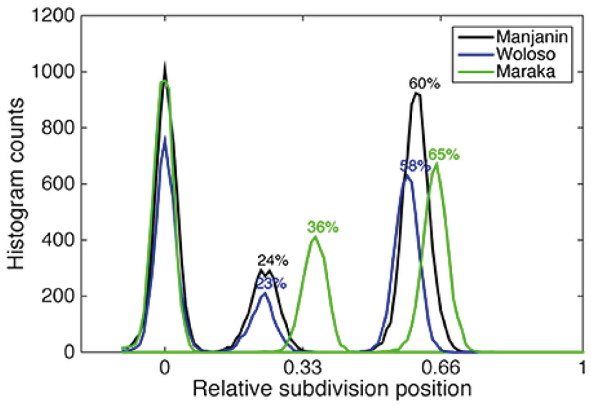



**FIGURE 3 |** Complex metrical structure found in djembe drumming from Mali. In this musical culture, while the top level of temporal organization is isochronous, the level below it is neither isochronous nor exhibits small integer ratios. For three musical pieces (Manjanin, Woloso, and Maraka), the figure shows the time of occurrence of other beats between two isochronous beats. Manjanin and Woloso are clearly anisochronous, while Maraka (green plot) is almost isochronous. Figure copied verbatim from Polak et al. (2016).

probability 1. This property makes isochronous sequences, from an information-theoretic perspective, purely deterministic and predictable, granting the lowest possible entropy. In other words, conditional on a known repetition rate (i.e., tempo), isochrony minimizes entropy in rhythmic sequences.

Physics offers prime examples of isochronous processes (Strogatz, 2000). For instance, atomic clocks are based on isochronous oscillations, i.e., atomic activity regularly occurring at known frequencies (Strogatz, 2003). Events like these, reliably repeating at regular time intervals, can serve as a benchmark (**Table 1**): they are used by mankind to synchronize their clocks and, for our purposes, they represent the highest level of isochrony achievable by a system. In other words, empirical isochrony can be defined as synchrony, i.e., perfect co-occurrence (Ravignani, 2017a), of an empirical sequence with respect to an isochronous reference grid.

Likewise, mathematical and computational models of isochrony are quite straightforward. In the simplest case, isochrony can be mathematically generated by trigonometric functions, such as sine and cosine (**Figure 5**). More realistic models of isochronous human behavior involve, for instance, long-range correlations (Madison and Delignieres, 2009), and fractal scaling (Madison, 2004). Isochronous synchrony between two or more entities can be modeled using phase resetting (Sismondo, 1990; Greenfield and Roizen, 1993), period bisection (Ravignani, 2014; Ravignani et al., 2014b), coupled oscillator models (Strogatz and Stewart, 1993; Large and Kolen, 1994; Strogatz, 2000; Rouse et al., 2016; Ravignani, 2017a), and a number of other techniques (reviewed in Ravignani and Norton, 2017). In dynamical systems, isochrony ranks amongst the best understood non-linear processes. The take-home message is that isochrony, in many of its forms and variants, can be comprehensively defined by very simple mathematical expressions (and visualized geometrically, see Toussaint, 2013; Ravignani, 2017b).

## PHYSIOLOGY AND NEUROSCIENCE

Human heartbeat, respiration, and locomotion all have an element of isochrony (**Figure 6**) in that they exhibit more regularity than random patterns but less regularity than periodic patterns in physical systems (Winfree, 1986). On the one hand, these processes can be quite regular within a short enough window of measurement (Larsson, 2012, 2014, 2015; Teie, 2016). In fact, the most commonly used measures of these physiological variables (beats, breaths or gaits per minute) assume *local isochrony* for the sequence analyzed. So, for example, most finger tapping studies have collected sequences of only 20–50 intervals in order to avoid the complicating drift (Wing and Kristofferson, 1973a,b; Madison, 2001). On the other hand, heartbeat, respiration and locomotion are highly dynamic and mutable in order to be functional (**Figure 6**); in other words, acceleration, deceleration, and phase shifts – all disrupting perfect isochrony – are quite common (Strogatz, 2003).

Isochrony and synchrony are also emergent properties of the nervous system. Synchronous groups of neurons, each oscillating isochronously, are common in the brain (Buzsaki,

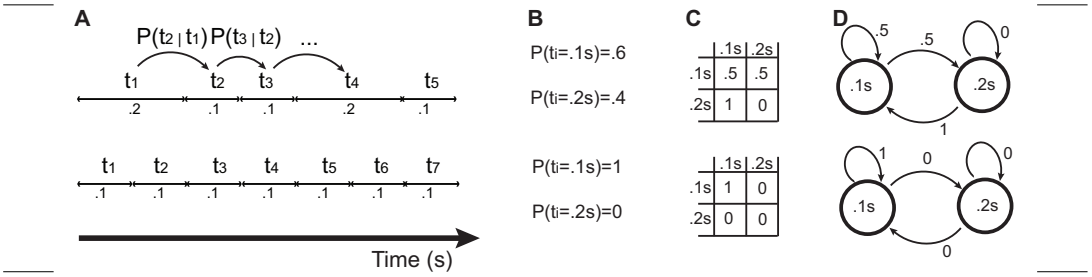



**FIGURE 4 |** Different ways of representing patterns of time intervals. The bottom shows an isochronous pattern, and the top a slightly more complex pattern. **(A)** Time series of intervals ($t_1$, $t_2$, ...), inducing transition probabilities $P(t_1 = x \mid t_2 = y)$, shortened as $P(t_1 \mid t_2)$. **(B)** Set of interval durations and probabilities of occurrence. **(C)** Transition matrices resulting from transition probabilities. **(D)** Equivalently, a probabilistic finite state machine also generates the patterns shown in **(A)** and described by the transition matrix in **(C)**.

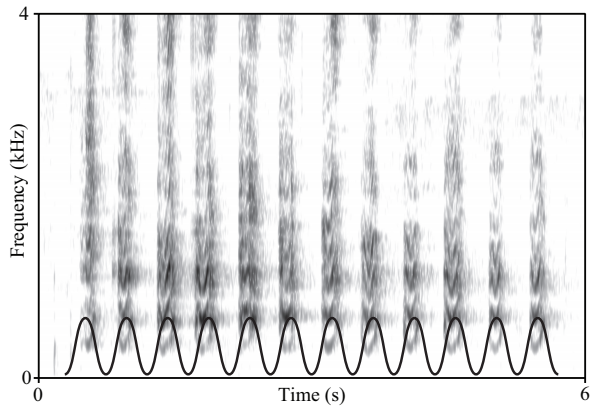



**FIGURE 5 |** Spectrogram of a California sea lion (*Zalophus californianus*) bark sequence, showing an extremely isochronous rhythm (Fourier window length = 0.02 s). To highlight this metronomic regularity, we have superimposed a sinusoidal wave to the lower part of the spectrogram. The sinusoid was plotted to feature as many maxima as the number of barks: the fact that each maximum aligns with one bark provides an intuitive qualification of the isochronicity of this pinniped vocalization. (Recording collected by AR with a Zoom H6 recorder and Zoom XY-6 microphone at 44.1 kHZ, 16-bit, at Santa Cruz pier, California, on October 2016).

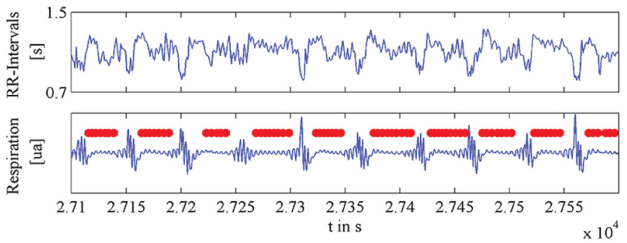



**FIGURE 6 |** Plot of exemplary time series of human heartbeat **(top)**, and respiration **(bottom)**, during and after apnea events. Figure modified from Penzel et al. (2016).

2006). At a higher level, cortico-subcortical networks are usually recruited to produce and perceive external isochronous events (Kotz and Schwartze, 2010, 2016; Nozaradan et al., 2011; Fujioka et al., 2012). Finally, some pathological states of the central nervous system are known to disrupt intentional isochrony, for instance Parkinson's disease (Grahn, 2012).

What is the neurophysiological basis for behavioral isochrony? Interesting connections between timing of vocalizations and neurophysiology have been discovered by physiologists working on non-human animals, for instance, amphibians. In some frog species, the temporal structure of courtship vocalizations is modulated by hormones (Zornik and Kelley, 2011). An outstanding question is, of course, whether these connections between temporal behavior and hormones can be found in humans, whose ethogram might be more complex than some amphibians. Recent findings in the human neurogenetics of music make this line of research quite promising (Granot et al., 2007; Ukkola et al., 2009; Ukkola-Vuoti et al., 2013; Kanduri et al., 2015a,b).

## COMPARATIVE COGNITION: NON-HUMAN ANIMALS

Some neural, behavioral and anatomical traits underlying isochronous rhythm perception and production are shared with a broad range of species (Wilson and Cook, 2016). These may either have a common evolutionary origin, or have evolved into similar traits under different evolutionary pressures (Ravignani et al., 2014b, 2016b). For instance, timing processes involving the basal ganglia and isochronous oscillations in the brain are shared with other primates and probably other animal taxa. Other traits are rare across species, only found in humans and a few other animals (Patel et al., 2009a,b). For instance, motor entrainment to an external isochronous pulse is only found sparingly in the animal kingdom (Fitch, 2015; Iversen, 2016; Ravignani et al., 2016b; Wilson and Cook, 2016).

A first, crucial difference separates human isochronous behavior from the examples of isochrony in nature provided above. This is the extent to which isochronous pattern production is driven and affected by external factors. A decaying isotope and a person walking at regular pace do not need an external oscillatory stimulus to keep producing isochronous behavior. These are cases of *endogenous* isochrony (Pikovsky et al., 2003), corresponding to self-sustained oscillators in physics. Conversely, the isochronous behavior exhibited by humans dancing or tapping to music is mostly *exogenous*: an internal pacemaker is partially corrected by externally perceived oscillatory activity (Merker et al., 2009), corresponding to forced or coupled oscillators in physics.

How about all shades of gray between these two extreme cases? Those animal species exhibiting isochrony are sparsely and heterogeneously divided among cases of endogenous and exogenous isochrony. This makes animal research key to understand the nature of human isochrony: for every particular type of isochrony found in a species, its neural mechanisms and resulting behaviors can be mapped and compared between that species and humans (Merchant et al., 2015).

Within exogenous isochrony, another distinction is between prediction and reaction (**Table 1**), depending on whether a timing event is produced by predicting when the next event should occur, or reacting to a previous event (Patel et al., 2009b). Humans exhibit exogenous *predictive* timing (Fujioka et al., 2012), while crickets exhibit exogenous *reactive* timing (Greenfield and Roizen, 1993). Finally, the isochronous tail wagging of dogs is quite likely to be endogenous, as external oscillatory stimuli are unlikely to affect its period or phase (cf. Buxton and Goodman, 1967; Fitch, 2009). While some species can be readily classified along the dimensions above (Rouse et al., 2016), for other species some data is available (**Figure 5**), though still not enough to be classified into isochrony types (e.g., Schusterman, 1977). Finally, for the majority of species, no systematic investigation of isochronous behavior has been performed. In other words, we still lack data on how most

species produce and perceive isochrony under a wide range of different conditions, which would be diagnostic to the underlying mechanisms and limitations.

In animal research, isochrony has been investigated using two main methods: observing natural behavior and training animals to produce specific temporal sequences. Isochrony as natural behavior in other animal species has long been studied, though its relevance to human rhythm has been pointed out only recently (Madison, 2004; Merker et al., 2009; Ravignani et al., 2014b). In fact, many animals signal over time in a precisely isochronous fashion (e.g., see **Figure 5**). 'Isochronous species' span crickets, frogs, fireflies, birds, crabs, and marine mammals (Schusterman, 1977; Sismondo, 1990; Greenfield and Roizen, 1993; Strogatz, 2003; Merker et al., 2009; Kahn et al., 2014; Norton and Scharff, 2016). As these studies are often purely behavioral and observational, rarely targeting neurobiological brain mechanisms, it is difficult to know whether isochrony is endogenous, exogenous, predictive or reactive.

An alternative is to test animals' capacities to produce isochrony in a controlled experimental setup. This is often done in conjunction with synchronization experiments. The animals are trained to produce specific isochronous behaviors, often with the purpose of entraining to a musical beat, and are then tested in their ability to generalize to different tempi and levels of jitter (Wilson and Cook, 2016). The only irrefutable results of *exogenous predictive isochrony* in any animal species are three: humans, a sea lion and a cockatoo (Patel et al., 2009a; Cook et al., 2013). Trained isochrony due instead to *reactive* timing might be more common: several species appear capable of producing series of temporal intervals of equal duration (Hasegawa et al., 2011; Hattori et al., 2015).

## ISOCHRONY IN INTERACTION

In human communication, two opposing functions affect the structure of the signal: expressivity and compressibility (**Table 1** and Kirby et al., 2015). Expressivity influences the amount of information content, hence semantics, which ideally should be maximized (Kirby et al., 2015). Compressibility refers to the density of information transmitted: intuitively, it is cost efficient to transmit the same amount of information in its shortest or maximally compressed form (MacKay, 2003). In other words, signalers would ideally broadcast the maximum quantity of information, using the least possible amount of signal. This tradeoff is further modulated by redundancy: a maximally compressible communication system with no redundancy can be irreversibly corrupted by a minimal transmission error. Hence, a signaler might not want to completely minimize entropy in order to leave room for redundancy.

To what extent can compressibility, expressivity, and redundancy account for human isochrony (**Figure 7**)? Mathematically, isochronous signals maximize redundancy and minimize entropy, but leave almost no room for expressivity. Comparatively, when human participants develop signaling systems in communication experiments, no expressivity leads to maximum compressibility (Kirby et al., 2008, 2015). This, by analogy, would dismiss isochronous pattern production as an expressive communication system (**Figure 7**), i.e., a system where signals are mapped to meanings (but see Bharucha and Pryor, 1986; Horr and Di Luca, 2015). However, the meaning of the transmitted message could lie in the signal emission *per se*, rather than that being broadcasted through the signal. This is the concept behind 'signaling signalhood' (Scott-Phillips et al., 2009): the message of an isochronous pattern is its 'isochronicity,' instead of being used in referential communication. Consider a hypothetical example. Take rhythmic sequences composed of only two durational intervals (such as the first two intervals of a sequence in **Figure 4A**). The set of all two sequences could be used for communicative purposes in two main ways. In the first, more common case of 'referential communication,' the duration of the two intervals could encode different conceptual properties. For instance, the first interval could be used to encode the size of a referent, while the second its brightness. Hence, a rhythm composed of a short and a long interval would communicate a small, dark object, while a long and a short interval would refer to a large bright object. This variability in the lengths of the intervals would grant expressivity. However, if isochronous sequences were the signals most frequently transmitted, this way of encoding signals could not be expressive, because all objects would end up being encoded as having an average size and brightness. In contrast, in the second case of 'signaling signalhood,' the two intervals would be used to communicate precision in isochronous pattern production, i.e., to signal isochronicity. Hence, a pattern as the bottom of **Figure 4A** would signal high precision in isochrony, while the top pattern in **Figure 4A** would signal poor isochrony. From this point of view, human perception and production of isochronous patterns might better fit the second, signaling signalhood framework, rather than the first, referential communication framework.

There is a close match between the most precise levels of isochrony that humans are capable of producing and those they are capable of perceiving (Madison and Merker, 2002; Merker et al., 2009). This match also offers some support for the hypothesis that isochrony might have been shaped for communicative purposes. In other words, a communication system, and in particular one that takes advantage of, and evolves from, perceptual biases (Ryan, 1998), will show a match between features of the signal and the capacities to perceive those features. For example, the plumages of many bird species reflect ultraviolet light, which humans and other species cannot see, while conspecific birds can readily perceive and use to select a mate (Andersson and Amundsen, 1997; Vorobyev et al., 1998; Eaton, 2005). We hypothesize that an analogous process might have resulted from isochrony (expanding on Merker, 1999, 2000), if this were a communicative trait. In particular, a communication system employed to transmit information about deviations from an isochronous pulse would evolve toward levels of precision comparable between production and perception (Merker, 1999, 2000). This comparable precision is exactly what can be observed in human motoric and perceptual isochrony (Madison and Merker, 2002; Merker et al., 2009), offering some preliminary, indirect support for a possible communicative function of isochrony.

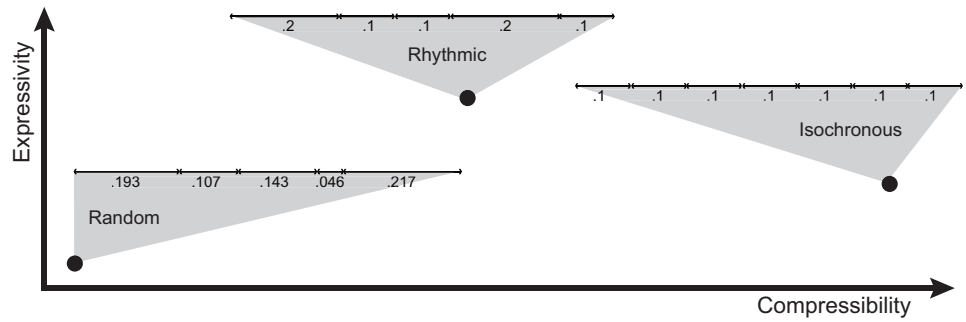



**FIGURE 7 |** Compressibility-expressivity tradeoff for temporal patterns. Three examples of timing patterns are provided, placing them in this compressibility-expressivity space. For each example, the black dot denotes the approximate location of the pattern. Each horizontal line represents one pattern, with the duration of each chunk expressed in seconds below it. Patterns used for communicative purposes are predicted to lie on or near the main diagonal, and patterns in the top-left and bottom-right quadrants of the diagram are hence not predicted to serve communicative purposes.

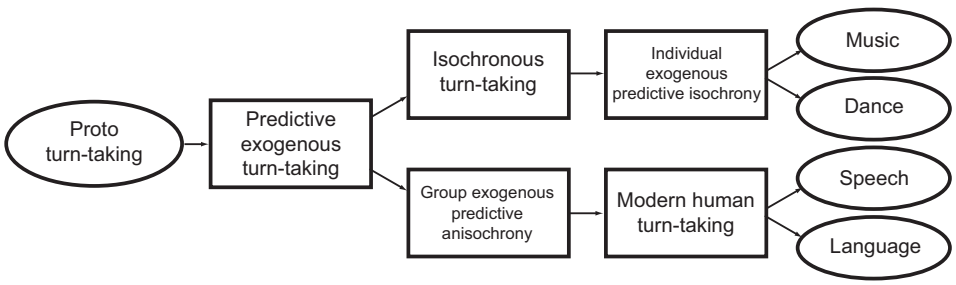



**FIGURE 8 |** Flux diagram of hypothetical precursors to isochronous synchrony and turn-taking. A hypothetical form of 'proto turn-taking' branches in two, leading to isochrony, as found in music and dance, and anisochronous turn-taking, as found in language and speech.

Isochrony does not appear to be used in the overt communication of modern humans, but might have played a role in some form of communication employed by our ancestors. In fact, isochrony is the optimal way to establish synchronized group signaling because it makes the duration of next interval perfectly predictable by another person or conspecific (Merker et al., 2009). This musical perspective on the evolution of isochrony connects to turn-taking, which is a crucial component of human language (**Figure 8**). Turn-taking allows speakers to effectively interact in conversation: it avoids that speakers' utterances overlap, while still enabling utterances to occur within a reasonable amount of time from each other. Interestingly, turn-taking in language is both predictive and exogenous, but seems to lack isochrony, except maybe in a few special cases. Still, turn-taking exhibits a particular temporal structure (Stivers et al., 2009; Levinson and Torreira, 2015). This structure appears to arise by a constant 200 ms lag (**Figure 9C**) between the ends and starts of utterances across cultures (Stivers et al., 2009), rather than a lag between the starts of consecutive utterances. This fixed-interval delay contrasts with the slightly positive or negative lags found in animal synchronization experiments (**Figures 9A,B**), and the anticipatory reaction in human musical synchronization. So, in modern humans, turn-taking is far from isochrony (except for when it is a by product of utterances having the same duration within and between speakers), but it might promote isochrony (Schultz et al., 2016). This makes turn-taking in modern organisms a potential approach to understand the evolution of isochrony (see **Figure 8**).

## EVOLUTIONARY HYPOTHESES AND FUTURE EMPIRICAL WORK

In conclusion, isochrony does not appear to relate to any *current* selection pressure. This is not surprising: a large number of evolved, heritable traits do not readily map to clear selection pressures in extant species. Instead, the pressures giving rise to isochrony might be sought in ancient humans, operating at some point between now and the split between our ancestor and that of chimpanzees/bonobos. The most articulate hypothesis to date proposes a multistage model (Merker, 1999, 2000; Merker et al., 2009). According to this, a recent ancestor to modern *Homo sapiens* would have been exposed to a selection pressure to attract migrating conspecific females. Accordingly, group vocalizations would have provided a conspicuously loud signal. The more individuals that managed to synchronize their calls, the greater the sound intensity, and the farther its reach in the terrain. In turn, the easiest way to achieve synchronization is to produce an isochronous signal, which is maximally predictable by the other callers, leading to an

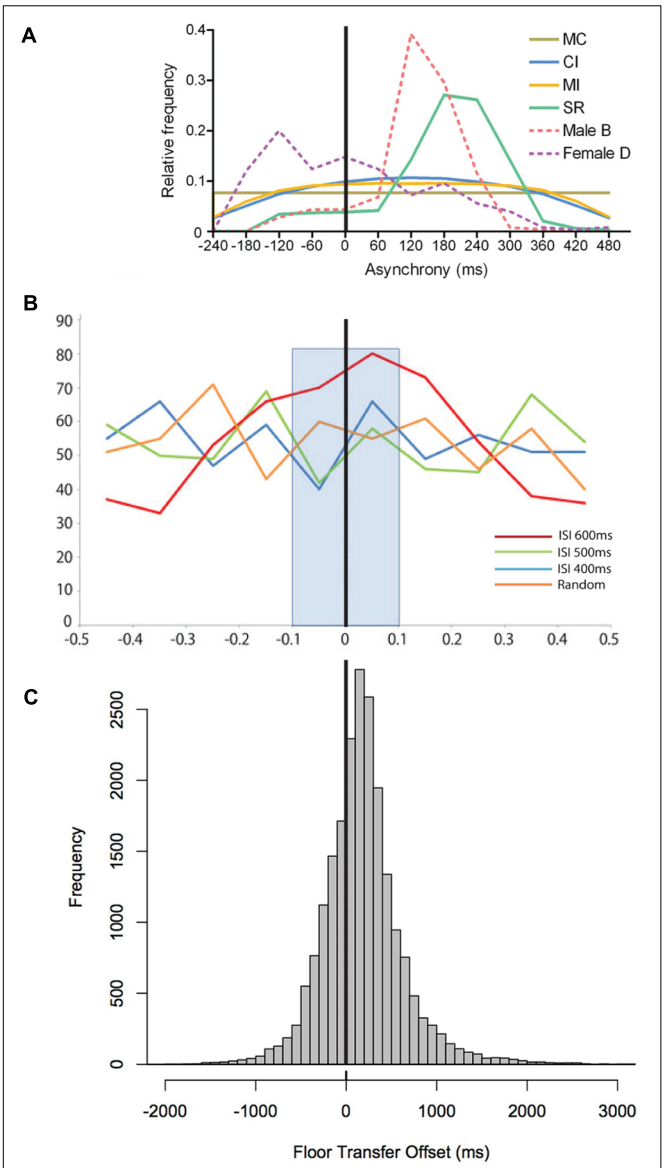



**FIGURE 9 |** Onset asynchronies in **(A)** a synchronization experiment in budgerigars (each bird, Male B and Female D, is a dashed line), **(B)** a tapping experiment in one individual chimpanzee, and **(C)** turn-taking in human conversation. Figure panels modified from Hasegawa et al. (2011) for **(A)**, Hattori et al. (2013) for **(B)**, and Levinson and Torreira (2015) for **(C)**.

isochronous, synchronous chorus (Merker et al., 2009). We further extend this idea, suggesting that early stages of vocal coordination could be precursors to two modern human traits: isochronous signaling in music and anisochronous turn-taking in language (**Figure 8**).

Our perspective focuses on the *function* of isochrony (Merker et al., 2009; Ravignani et al., 2014b), rather than its underlying ontogeny and phylogeny (Tinbergen, 1963). Hence, most ideas presented here are in principle compatible with other hypotheses focused on developmental trajectories or mechanisms (Iversen, 2016; Teie, 2016). The functional reason why isochrony appeared in human evolution might therefore not be directly derivable from how isochrony appears in development, or from its neurobehavioral mechanisms. The mechanisms underlying isochrony which manifest through development (Eerola et al., 2006) might also be reflected in evolution, though they do not need to be (see Gould, 1977, and subsequent debate).

Research on modern humans provides some grounds, in principle, for isochrony to have been an evolutionary selected trait. In particular, isochronous timing seems quite variable across individuals, and enhanced by learning (Max and Yudman, 2003; Manning and Schutz, 2016; Tierney et al., 2017). Individual differences and learning plasticity are neither necessary nor sufficient conditions to show that a trait, such as isochrony, underwent evolution by natural selection. However, individual differences are often a prerequisite for natural selection to act on a trait. Likewise, learning plasticity can be an outcome of an evolutionary process acting on behavior and cognition, rather than on a physical trait.

For the purpose of the present paper, it would be interesting to find the genetic and neuro-hormonal biological substrates responsible for perception and production of isochronous behavior both in humans and other animals. In fact, as isochrony appears as a relatively simple behavioral trait, study of its neuro-genetic and hormonal substrates might prove an initial building block to understand rhythm more in general.

Our suggestions can be readily tested along several strands of empirical research in humans. Temporal interactions in groups of animals are known to lead to isochrony as one of the equilibrium outcomes (Sismondo, 1990; Greenfield and Roizen, 1993; Kahn et al., 2014). Human data on turn-taking may be re-analyzed, asking whether the isochrony of each partner entails the best predictability of turn-taking vs. constant-lag alternation. Likewise, the large body of research on isochrony perception and production across modalities and domains may be synthetized (Iversen et al., 2015; Celma-Miralles et al., 2016), to examine (1) the limits and boundaries of the human sense of isochrony, and (2) which experiments are lacking that would entail comparability across domains and modalities. In general, the field of rhythm would benefit by a tighter connection between individual and group processes: individual behavioral traits do not evolve in a vacuum, and individual timing might be modulated by social factors (Ravignani et al., 2014a,b; Ravignani and Cook, 2016; Schirmer et al., 2016). For instance, in some primate and avian species, singing is accurately timed with the group depending on the sex and social status of each individual singer (Mann et al., 2006; Gamba et al., 2016). With this logic in mind, we can list a number of specific outstanding questions:

- Do individual and group isochrony influence each other (Ravignani et al., 2014b)? How are they related at a mechanistic, functional, developmental and phylogenetic level (Kirschner and Tomasello, 2009; Ravignani et al., 2016b)?
- What are the effects of training, conformity, and social cues on the perception and production of isochronous behaviors (Schirmer et al., 2016)?

- How do different individual group interaction modes, for instance coordination and competition, map on to temporal patterns produced, such as isochrony (Ravignani et al., 2014a)?

These questions are not only relevant from the perspective of human cognitive neuroscience and animal behavior. They are also key to test evolutionary hypotheses, where the fitness landscape might be influenced by social factors and cultural niches (Tomasello, 2009; Boyd et al., 2011; Kendal, 2011).

Finally, while selection pressures in our ancestors are difficult to reconstruct, their effects might still be observable in the behavioral tendencies, genome and neuroendocrine system of modern humans (Holmquist and Vestin, 2010; Madison, 2011; Björk, 2013; Madison et al., 2017, in press). For instance, recent studies have mapped musical and rhythmic phenotypes to genes and hormonal profiles (Mosing et al., 2015; Miani, 2016a,b); more focused studies linking biology and psychology are needed for the specific trait(s) underlying isochrony.

## AUTHOR CONTRIBUTIONS

AR wrote a first draft of the manuscript. GM provided references, guidance, and advice, and edited the manuscript.

## FUNDING

AR has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 665501 with the research Foundation Flanders (FWO) (Pegasus$^2$ Marie Curie fellowship 12N5517N awarded to AR), a visiting fellowship in Language Evolution from the Max Planck Society (awarded to AR), and ERC grant 283435 ABACUS (awarded to Bart de Boer).

## ACKNOWLEDGMENT

AR is grateful to Steve Levinson for advice and support, and to Bill Thompson and Heikki Rasilo for the stimulating discussions in Tennessee, which inspired this paper.

## REFERENCES

Andersson, S., and Amundsen, T. (1997). Ultraviolet colour vision and ornamentation in bluethroats. *Proc. R. Soc. Lond. B Biol. Sci.* 264, 1587–1591. doi: 10.1098/rspb.1997.0221

Arnal, L. H., and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390–398. doi: 10.1016/j.tics.2012.05.003

Arom, S. (1991). *African Polyphony and Polyrhythm: Musical Structure and Methodology*, Trans. M. Thom, B. Tuckett, and R. Boyd (Cambridge: Cambridge university press). doi: 10.1017/CBO9780511518317

Aubanel, V., Davis, C., and Kim, J. (2016). Exploring the role of brain oscillations in speech perception in noise: intelligibility of isochronously retimed speech. *Front. Hum. Neurosci.* 10:430. doi: 10.3389/fnhum.2016.00430

Bégel, V., Benoit, C.-E., Correa, A., Cutanda, D., Kotz, S. A., and Dalla Bella, S. (2017). "Lost in time" but still moving to the beat. *Neuropsychologia* 94, 129–138. doi: 10.1016/j.neuropsychologia.2016.11.022

Bharucha, J. J., and Pryor, J. H. (1986). Disrupting the isochrony underlying rhythm: an asymmetry in discrimination. *Percept. Psychophys.* 40, 137–141. doi: 10.3758/BF03203008

Björk, J. (2013). *Could Musical Mastery Affect how Attractive A Person is Rated As A Prospective Partner?*. B.Sc. thesis, Umeå University, Umeå.

Bolton, T. L. (1894). Rhythm. *Am. J. Psychol.* 6, 145–238. doi: 10.2307/1410948

Bowling, D. L., Herbst, C. T., and Fitch, W. T. (2013). Social origins of rhythm? synchrony and temporal regularity in human vocalization. *PLOS ONE* 8:e80402. doi: 10.1371/journal.pone.0080402

Bowling, D. L., Hoeschele, M., Gill, K. Z., and Fitch, W. T. (2017). The nature and nurture of musical consonance. *Music Percept.* 35, 118–121. doi: 10.1525/mp.2017.35.1.118

Boyd, R., Richerson, P. J., and Henrich, J. (2011). The cultural niche: why social learning is essential for human adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 2), 10918–10925. doi: 10.1073/pnas.1100290108

Brown, S., Pfordresher, P. Q., and Chow, I. (2017). A musical model of speech rhythm. *Psychomusicology* 27, 95. doi: 10.1037/pmu0000175

Buxton, D. F., and Goodman, D. C. (1967). Motor function and the corticospinal tracts in the dog and raccoon. *J. Comp. Neurol.* 129, 341–360. doi: 10.1002/cne.901290405

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195301069.001.0001

Celma-Miralles, A., de Menezes, R. F., and Toro, J. M. (2016). Look at the beat, feel the meter: top–down effects of meter induction on auditory and visual modalities. *Front. Hum. Neurosci.* 10:108. doi: 10.3389/fnhum.2016.00108

Cohen, J. E. (1962). Information theory and music. *Behav. Sci.* 7, 137–163. doi: 10.1002/bs.3830070202

Cook, P., Rouse, A., Wilson, M., and Reichmuth, C. J. (2013). A California sea lion (*Zalophus californianus*) can keep the beat: motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *J. Comp. Psychol.* 127, 1–16. doi: 10.1037/a0032345

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *J. Phonet.* 11, 51–62.

Eaton, M. D. (2005). Human vision fails to distinguish widespread sexual dichromatism among sexually "monochromatic" birds. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10942–10946. doi: 10.1073/pnas.0501891102

Eerola, T., Luck, G., and Toiviainen, P. (2006). An investigation of pre-schoolers' corporeal synchronization with music. *Paper Presented at the Proceedings of the 9th International Conference on Music Perception and Cognition*, Bologna.

Fabb, N., and Halle, M. (2012). "Grouping in the stressing of words, in metrical verse, and in music," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohmeier, J. A. Hawkins, and I. Cross (Oxford: Oxford University Press), 4–21.

Fink, B., and Shackelford, T. K. (2016). Why did dance evolve? a comment on Laland, Wilkins, and Clayton (2016). *Evol. Psychol. Sci.* 3, 147–148. doi: 10.1007/s40806-016-0075-3

Fitch, W. T. (2009). *The Biology and Evolution of Rhythm: Unraveling a Paradox Language and Music as Cognitive Systems*. Oxford: Oxford University Press.

Fitch, W. T. (2015). "The biology and evolution of musical rhythm: an update," in *Structures in the Mind: Essays on Language, Music, and Cognition in Honor of Ray Jackendoff*, eds D. I. Toivonen, P. Csúri, and E. van der Zee (Cambridge, MA: MIT Press), 293–324.

Fitch, W. T. (2016). Dance, music, meter and groove: a forgotten partnership. *Front. Hum. Neurosci.* 10:64. doi: 10.3389/fnhum.2016.00064

Fitch, W. T. (2017). Cultural evolution: lab-cultured musical universals. *Nat. Hum. Behav.* 1, 1–2. doi: 10.1038/s41562-016-0018

Friberg, A., and Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *J. Acoust. Soc. Am.* 98, 2524–2531. doi: 10.1121/1.413218

Fujii, S., and Wan, C. Y. (2014). The role of rhythm in speech and language rehabilitation: the SEP hypothesis. *Front. Hum. Neurosci.* 8:777. doi: 10.3389/fnhum.2014.00777

Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *J. Neurosci.* 32, 1791–1802. doi: 10.1523/JNEUROSCI.4107-11.2012

Gamba, M., Torti, V., Estienne, V., Randrianarison, R. M., Valente, D., Rovara, P., et al. (2016). The indris have got rhythm! Timing and pitch variation of a primate song examined between sexes and age classes. *Front. Neurosci.* 10:249. doi: 10.3389/fnins.2016.00249

Gould, S. J. (1977). *Ontogeny and Phylogeny*. Cambridge, MA: Harvard University Press.

Grabe, E., and Low, E. L. (2002). "Durational variability in speech and the rhythm class hypothesis," in *Papers in Laboratory Phonology*, 7, eds N. Warner and C. Gussenhoven (Berlin: Mouton de Gruyter), 515–546.

Grahn, J. A. (2012). Neural mechanisms of rhythm perception: current findings and future perspectives. *Top. Cogn. Sci.* 4, 585–606. doi: 10.1111/j.1756-8765.2012.01213.x

Granot, R. Y., Frankel, Y., Gritsenko, V., Lerer, E., Gritsenko, I., Bachner-Melman, R., et al. (2007). Provisional evidence that the arginine vasopressin 1a receptor gene is associated with musical memory. *Evol. Hum. Behav.* 28, 313–318. doi: 10.1016/j.evolhumbehav.2007.05.003

Greenfield, M. D., and Roizen, I. (1993). Katydid synchronous chorusing is an evolutionarily stable outcome of female choice. *Nature* 364, 618–620. doi: 10.1038/364618a0

Hasegawa, A., Okanoya, K., Hasegawa, T., and Seki, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Sci. Rep.* 1:120. doi: 10.1038/srep00120

Hattori, Y., Tomonaga, M., and Matsuzawa, T. (2013). Spontaneous synchronized tapping to an auditory rhythm in a chimpanzee. *Sci. Rep.* 3:1566. doi: 10.1038/srep01566

Hattori, Y., Tomonaga, M., and Matsuzawa, T. (2015). Distractor effect of auditory rhythms on self-paced tapping in chimpanzees and humans. *PLOS ONE* 10:e0130682. doi: 10.1371/journal.pone.0130682

Holmquist, J., and Vestin, M. (2010). *Will Musical Skill Attract Mates? An Experimental Study of Mate Preferences as a Function of Musical Ability*. M.Sc. thesis, Umeå University, Umeå.

Honing, H., Ladinig, O., Háden, G. P., and Winkler, I. (2009). Is beat induction innate or learned? *Ann. N. Y. Acad. Sci.* 1169, 93–96. doi: 10.1111/j.1749-6632.2009.04761.x

Horr, N. K., and Di Luca, M. (2015). Taking a long look at isochrony: perceived duration increases with temporal, but not stimulus regularity. *Atten. Percept. Psychophys.* 77, 592–602. doi: 10.3758/s13414-014-0787-z

Iversen, J. R. (2016). "In the beginning was the beat: evolutionary origins of musical rhythm in humans," in *The Cambridge Companion to Percussion*, ed. R. Hartenberger (Cambridge: Cambridge University Press).

Iversen, J. R., Patel, A. D., Nicodemus, B., and Emmorey, K. (2015). Synchronization to auditory and visual rhythms in hearing and deaf individuals. *Cognition* 134, 232–244. doi: 10.1016/j.cognition.2014.10.018

Jacoby, N., and McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr. Biol.* 27, 359–370. doi: 10.1016/j.cub.2016.12.031

Jadoul, Y., Ravignani, A., Thompson, B., Filippi, P., and de Boer, B. (2016). Seeking temporal predictability in speech: comparing statistical approaches on 18 world languages. *Front. Hum. Neurosci.* 10:586. doi: 10.3389/fnhum.2016.00586

Kahn, A. T., Holman, L., and Backwell, P. R. (2014). Female preferences for timing in a fiddler crab with synchronous courtship waving displays. *Anim. Behav.* 98, 35–39. doi: 10.1016/j.anbehav.2014.09.028

Kanduri, C., Kuusi, T., Ahvenainen, M., Philips, A. K., Lähdesmäki, H., and Järvelä, I. (2015a). The effect of music performance on the transcriptome of professional musicians. *Sci. Rep.* 5:9506. doi: 10.1038/srep09506

Kanduri, C., Raijas, P., Ahvenainen, M., Philips, A. K., Ukkola-Vuoti, L., Lähdesmäki, H., et al. (2015b). The effect of listening to music on human transcriptome. *PeerJ* 3:e830. doi: 10.7717/peerj.830

Kendal, J. R. (2011). Cultural niche construction and human learning environments: Investigating sociocultural perspectives. *Biol. Theory* 6, 241–250. doi: 10.1007/s13752-012-0038-2

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10681–10686. doi: 10.1073/pnas.0707835105

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141, 87–102. doi: 10.1016/j.cognition.2015.03.016

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. doi: 10.1016/j.jecp.2008.07.005

Kotz, S. A., and Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn. Sci.* 14, 392–399. doi: 10.1016/j.tics.2010.06.005

Kotz, S. A., and Schwartze, M. (2016). "Motor-timing and sequencing in speech production: a general-purpose framework," in *Neurobiology of Language*, eds G. Hickok and S. L. Small (San Diego, CA: Academic Press), 717–724. doi: 10.1016/B978-0-12-407794-2.00057-2

Laland, K., Wilkins, C., and Clayton, N. (2016). The evolution of dance. *Curr. Biol.* 26, R5–R9. doi: 10.1016/j.cub.2015.11.031

Large, E. W., and Kolen, J. F. (1994). Resonance and the perception of musical meter. *Conn. Sci.* 6, 177–208. doi: 10.1080/09540099408915723

Larsson, M. (2012). Incidental sounds of locomotion in animal cognition. *Anim. Cogn.* 15, 1–13. doi: 10.1007/s10071-011-0433-2

Larsson, M. (2014). Self-generated sounds of locomotion and ventilation and the evolution of human rhythmic abilities. *Anim. Cogn.* 17, 1–14. doi: 10.1007/s10071-013-0678-z

Larsson, M. (2015). Tool-use-associated sound in the evolution of language. *Anim. Cogn.* 18, 993–1005. doi: 10.1007/s10071-015-0885-x

Launay, J., Grube, M., and Stewart, L. (2014). Dysrhythmia: a specific congenital rhythm perception deficit. *Front. Psychol.* 5:18. doi: 10.3389/fpsyg.2014.00018

Lehiste, I. (1977). Isochrony reconsidered. *J. Phonet.* 5, 253–263.

Levinson, S. C., and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Front. Psychol.* 6:731. doi: 10.3389/fpsyg.2015.00731

MacKay, D. J. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.

Madison, G. (2001). Variability in isochronous tapping: higher order dependencies as a function of intertap interval. *J. Exp. Psychol.* 27, 411–422. doi: 10.1037/0096-1523.27.2.411

Madison, G. (2004). Fractal modeling of human isochronous serial interval production. *Biol. Cybern.* 90, 105–112. doi: 10.1007/s00422-003-0453-3

Madison, G. (2009). An auditory illusion of infinite tempo change based on multiple temporal levels. *PLOS ONE* 4:e8151. doi: 10.1371/journal.pone.0008151

Madison, G. (2011). "Cause and affect: a functional perspective on music and emotion," in *Art and the Senses*, eds F. Bacci and D. Melcher (Oxford: Oxford University Press), 329–350.

Madison, G. (2014). Sensori-motor synchronisation variability decreases as the number of metrical levels in the stimulus signal increases. *Acta Psychol.* 147, 10–16. doi: 10.1016/j.actpsy.2013.10.002

Madison, G., and Delignieres, D. (2009). Auditory feedback affects the long-range correlation of isochronous serial interval production: support for a closed-loop or memory model of timing. *Exp. Brain Res.* 193, 519–527. doi: 10.1007/s00221-008-1652-x

Madison, G., Holmquist, J., and Vestin, M. (in press). Musical improvisation skill in a prospective partner is associated with mate value and preferences, consistent with sexual selection and parental investment theory: implications for the origin of music. *Evol. Hum. Behav.*

Madison, G., and Merker, B. (2002). On the limits of anisochrony in pulse attribution. *Psychol. Res.* 66, 201–207. doi: 10.1007/s00426-001-0085-y

Madison, G., and Merker, B. (2004). Human sensorimotor tracking of continuous subliminal deviations from isochrony. *Neurosci. Lett.* 370, 69–73. doi: 10.1016/j.neulet.2004.07.094

Madison, G., Ullén, F., and Merker, B. (2017). "Metrically structured time and entrainment," in *The Routledge Companion to Embodied Music Interaction*, eds M. Lesaffre, M. Leman, and P. Maes (London: Routledge).

Mann, N. I., Dingess, K. A., and Slater, P. (2006). Antiphonal four-part synchronized chorusing in a Neotropical wren. *Biol. Lett.* 2, 1–4. doi: 10.1098/rsbl.2005.0373

Manning, F. C., and Schutz, M. (2016). Trained to keep a beat: movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychol. Res.* 80, 532–542. doi: 10.1007/s00426-015-0678-5

Max, L., and Yudman, E. M. (2003). Accuracy and variability of isochronous rhythmic timing across motor systems in stuttering versus nonstuttering individuals. *J. Speech Lang. Hear. Res.* 46, 146–163. doi: 10.1044/1092-4388(2003/012)

McAuley, J. D. (2010). *Tempo and Rhythm Music Perception*. Berlin: Springer, 165–199. doi: 10.1007/978-1-4419-6114-3_6

Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., and Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. B* 370:20140093. doi: 10.1098/rstb.2014.0093

Merker, B. (1999). Synchronous chorusing and the origins of music. *Music. Sci.* 3, 59–73. doi: 10.1177/10298649000030S105

Merker, B. (2000). "Synchronous chorusing and human origins," in *The Origins of Music*, eds N. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 315–327.

Merker, B. (2014). Groove or swing as distributed rhythmic consonance: introducing the groove matrix. *Front. Hum. Neurosci.* 8:454. doi: 10.3389/fnhum.2014.00454

Merker, B., Madison, G., and Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex* 45, 4–17. doi: 10.1016/j.cortex.2008.06.011

Miani, A. (2016a). *The Impact of Sex and Menstrual Cycle on Rhythmic Synchronization: An Experimental Approach*. Master thesis, Aarhus University, Aarhus.

Miani, A. (2016b). Sexual arousal and rhythmic synchronization: a possible effect of vasopressin. *Med. Hypotheses* 93, 122–125. doi: 10.1016/j.mehy.2016.05.030

Mosing, M. A., Verweij, K. J., Madison, G., Pedersen, N. L., Zietsch, B. P., and Ullén, F. (2015). Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evol. Hum. Behav.* 36, 359–366. doi: 10.1016/j.evolhumbehav.2015.02.004

Motz, B. A., Erickson, M. A., and Hetrick, W. P. (2013). To the beat of your own drum: cortical regularization of non-integer ratio rhythms toward metrical patterns. *Brain Cogn.* 81, 329–336. doi: 10.1016/j.bandc.2013.01.005

Neuhoff, H., Polak, R., and Fischinger, T. (2017). Perception and evaluation of timing patterns in drum ensemble music from mali. *Music Percept.* 34, 438–451. doi: 10.1525/mp.2017.34.4.438

Norton, P., and Scharff, C. (2016). "Bird song metronomics": isochronous organization of zebra finch song rhythm. *Front. Neurosci.* 10:309. doi: 10.3389/fnins.2016.00309

Nozaradan, S., Peretz, I., Missal, M., and Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* 31, 10234–10240. doi: 10.1523/JNEUROSCI.0411-11.2011

Pagliarini, E., Scocchia, L., Vernice, M., Zoppello, M., Balottin, U., Bouamama, S., et al. (2017). Children's first handwriting productions show a rhythmic structure. *Sci. Rep.* 7:5516. doi: 10.1038/s41598-017-05105-6

Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009a). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830. doi: 10.1016/j.cub.2009.03.038

Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009b). Studying synchronization to a musical beat in nonhuman animals. *Ann. N. Y. Acad. Sci.* 1169, 459–469. doi: 10.1111/j.1749-6632.2009.04581.x

Penzel, T., Kantelhardt, J. W., Bartsch, R. P., Riedl, M., Kraemer, J. F., Wessel, N., et al. (2016). Modulations of heart rate, ECG, and cardio-respiratory coupling observed in polysomnography. *Front. Physiol.* 7:460. doi: 10.3389/fphys.2016.00460

Pikovsky, A., Rosenblum, M., and Kurths, J. (2003). *Synchronization: A Universal Concept in Nonlinear Sciences*, Vol. 12. Cambridge: Cambridge University Press.

Polak, R., London, J., and Jacoby, N. (2016). Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: a corpus study of malian jembe drumming. *Front. Neurosci.* 10:285. doi: 10.3389/fnins.2016.00285

Ravignani, A. (2014). Chronometry for the chorusing herd: Hamilton's legacy on context-dependent acoustic signalling-a comment on Herbers (2013). *Biol. Lett.* 10, 20131018. doi: 10.1098/rsbl.2013.1018

Ravignani, A. (2017a). Interdisciplinary debate: agree on definitions of synchrony. *Nature* 545, 158. doi: 10.1038/545158c

Ravignani, A. (2017b). Visualizing and interpreting rhythmic patterns using phase space plots. *Music Percept.* 34, 557–568. doi: 10.1525/mp.2017.34.5.557

Ravignani, A., Bowling, D., and Kirby, S. (2014a). The psychology of biological clocks: a new framework for the evolution of rhythm. *Paper Presented at the Evolution of Language: Proceedings of the* 10th international conference, Singapore.

Ravignani, A., Bowling, D. L., and Fitch, W. T. (2014b). Chorusing, synchrony and the evolutionary functions of rhythm. *Front. Psychol.* 5:1118. doi: 10.3389/fpsyg.2014.01118

Ravignani, A., and Cook, P. (2016). The evolutionary biology of dance without frills. *Curr. Biol.* 26, R878–R879. doi: 10.1016/j.cub.2016.07.076

Ravignani, A., Delgado, T., and Kirby, S. (2016a). Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* 1:0007.

Ravignani, A., Fitch, W. T., Hanke, F. D., Heinrich, T., Hurgitsch, B., Kotz, S. A., et al. (2016b). What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Front. Neurosci.* 10:274. doi: 10.3389/fnins.2016.00274

Ravignani, A., Honing, H., and Kotz, S. A. (2017). The evolution of rhythm cognition: timing in music and speech. *Front. Hum. Neurosci.* 11:303. doi: 10.3389/fnhum.2017.00303

Ravignani, A., and Norton, P. (2017). Measuring rhythmic complexity: a primer to quantify and compare temporal structure in speech, movement, and animal vocalizations. *J. Lang. Evol.* 2, 4–19. doi: 10.1093/jole/lzx002

Repp, B. H. (2005). Sensorimotor synchronization: a review of the tapping literature. *Psychon. Bullet. Rev.* 12, 969–992. doi: 10.3758/BF03206433

Repp, B. H., and Su, Y.-H. (2013). Sensorimotor synchronization: a review of recent research (2006-2012). *Psychon. Bullet. Rev.* 20, 403–452. doi: 10.3758/s13423-012-0371-2

Richter, J., and Ostovar, R. (2016). "It don't mean a thing if it ain't got that Swing"–an alternative concept for understanding the evolution of dance and music in human beings. *Front. Hum. Neurosci.* 10:485. doi: 10.3389/fnhum.2016.00485

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed'languages. *Linguist. Controversies* 73, 79.

Rouse, A. A., Cook, P. F., Large, E. W., and Reichmuth, C. (2016). Beat keeping in a sea lion as coupled oscillation: implications for comparative understanding of human rhythm. *Front. Neurosci.* 10:257. doi: 10.3389/fnins.2016.00257

Ryan, M. J. (1998). Sexual selection, receiver biases, and the evolution of sex differences. *Science* 281, 1999–2003. doi: 10.1126/science.281.5385.1999

Schirmer, A., Meck, W. H., and Penney, T. B. (2016). The socio-temporal brain: connecting people in time. *Trends Cogn. Sci.* 20, 760–772. doi: 10.1016/j.tics.2016.08.002

Schultz, B. G., O'Brien, I., Phillips, N., McFarland, D. H., Titone, D., and Palmer, C. (2016). Speech rates converge in scripted turn-taking conversations. *Appl. Psycholinguist.* 37, 1201–1220. doi: 10.1017/S0142716415000545

Schusterman, R. J. (1977). Temporal patterning in sea lion barking (*Zalophus californianus*). *Behav. Biol.* 20, 404–408. doi: 10.1016/S0091-6773(77)90964-6

Scott-Phillips, T. C., Kirby, S., and Ritchie, G. R. (2009). Signalling signalhood and the emergence of communication. *Cognition* 113, 226–233. doi: 10.1016/j.cognition.2009.08.009

Sismondo, E. (1990). Synchronous, alternating, and phase-locked stridulation by a tropical katydid. *Science* 249, 55–58. doi: 10.1126/science.249.4964.55

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10587–10592. doi: 10.1073/pnas.0903616106

Strogatz, S. H. (2000). From Kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* 143, 1–20. doi: 10.1016/S0167-2789(00)00094-4

Strogatz, S. H. (2003). *Sync: The Emerging Science of Spontaneous Order*. New York City, NY: Hyperion Books.

Strogatz, S. H., and Stewart, I. (1993). Coupled oscillators and biological synchronization. *Sci. Am.* 269, 102–109. doi: 10.1038/scientificamerican1293-102

Su, Y.-H. (2016a). Sensorimotor synchronization with different metrical levels of point-light dance movements. *Front. Hum. Neurosci.* 10:186. doi: 10.3389/fnhum.2016.00186

Su, Y.-H. (2016b). Visual enhancement of illusory phenomenal accents in non-isochronous auditory rhythms. *PLOS ONE* 11:e0166880. doi: 10.1371/journal.pone.0166880

Teie, D. (2016). A comparative analysis of the universal elements of music and the fetal environment. *Front. Psychol.* 7:1158. doi: 10.3389/fpsyg.2016.01158

Tierney, A., White-Schwoch, T., MacLean, J., and Kraus, N. (2017). Individual differences in rhythmic skills: links with neural consistency and linguistic ability. *J. Cogn. Neurosci.* 29, 855–868. doi: 10.1162/jocn_a_01092

Tinbergen, N. (1963). On aims and methods of ethology. *Z. Tierpsychol.* 20, 410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x

Tomasello, M. (2009). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Toussaint, G. T. (2013). *The Geometry of Musical Rhythm: What Makes a" good" Rhythm Good?*. Boca Raton, FL: CRC Press.

Tuller, B., and Fowler, C. A. (1980). Some articulatory correlates of perceptual isochrony. *Atten. Percept. Psychophys.* 27, 277–283. doi: 10.3758/BF03206115

Ukkola, L. T., Onkamo, P., Raijas, P., Karma, K., and Järvelä, I. (2009). Musical aptitude is associated with AVPR1A-haplotypes. *PLOS ONE* 4:e5534. doi: 10.1371/journal.pone.0005534

Ukkola-Vuoti, L., Kanduri, C., Oikkonen, J., Buck, G., Blancher, C., Raijas, P., et al. (2013). Genome-wide copy number variation analysis in extended families and unrelated individuals characterized for musical aptitude and creativity in music. *PLOS ONE* 8:e56356. doi: 10.1371/journal.pone.0056356

Vorobyev, M., Osorio, D., Bennett, A. T., Marshall, N., and Cuthill, I. (1998). Tetrachromacy, oil droplets and bird plumage colours. *J. Comp. Physiol. A* 183, 621–633. doi: 10.1007/s003590050286

Wade, B. C. (2004). *Thinking Musically: Experiencing Music, Expressing Culture*. New York, NY: Oxford University Press.

Wilson, M., and Cook, P. F. (2016). Rhythmic entrainment: why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychon. Bullet. Rev.* 23, 1647–1659. doi: 10.3758/s13423-016-1013-x

Winfree, A. T. (1986). *Timing of Biological Clocks*. New York City, NY: Henry Holt and Company.

Wing, A. M., and Kristofferson, A. B. (1973a). Response delays and the timing of discrete motor responses. *Atten. Percept. Psychophys.* 14, 5–12.

Wing, A. M., and Kristofferson, A. B. (1973b). The timing of interresponse intervals. *Atten. Percept. Psychophys.* 13, 455–460. doi: 10.3758/BF03205802

Zornik, E., and Kelley, D. B. (2011). A neuroendocrine basis for the hierarchical control of frog courtship vocalizations. *Front. Neuroendocrinol.* 32:353–366. doi: 10.1016/j.yfrne.2010.12.006

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2017 Ravignani and Madison. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Neuroscience

MINI REVIEW
published: 06 February 2018
doi: 10.3389/fnins.2018.00020

# The Evolution of Musicality: What Can Be Learned from Language Evolution Research?

*Andrea Ravignani*[1,2,3]*, *Bill Thompson*[1,2] *and Piera Filippi*[1,4,5,6]

[1] Department of Language and Cognition, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [2] Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium, [3] Research Department, Sealcentre Pieterburen, Pieterburen, Netherlands, [4] Institute of Language, Communication and the Brain, Aix-en-Provence, France, [5] Laboratoire Parole et Langage LPL UMR 7309, Centre National de la Recherche Scientifique, Aix-Marseille Université, Aix-en-Provence, France, [6] Laboratoire de Psychologie Cognitive LPC UMR7290, Centre National de la Recherche Scientifique, Aix-Marseille Université, Marseille, France

Language and music share many commonalities, both as natural phenomena and as subjects of intellectual inquiry. Rather than exhaustively reviewing these connections, we focus on potential cross-pollination of methodological inquiries and attitudes. We highlight areas in which scholarship on the evolution of language may inform the evolution of music. We focus on the value of coupled empirical and formal methodologies, and on the futility of *mysterianism,* the declining view that the nature, origins and evolution of language cannot be addressed empirically. We identify key areas in which the evolution of language as a discipline has flourished historically, and suggest ways in which these advances can be integrated into the study of the evolution of music.

**Keywords: evolution of music, evolution of language, cultural transmission, cultural evolution, universals, music cognition, comparative cognition, nature and nurture**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky,*
*Independent Researcher,*
*Los Angeles, CA, United States*

***Reviewed by:***
*Mark Reybrouck,*
*KU Leuven, Belgium*
*Steven Brown,*
*McMaster University, Canada*

***Correspondence:**
*Andrea Ravignani*
*andrea.ravignani@gmail.com*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *15 September 2017*
***Accepted:*** *10 January 2018*
***Published:*** *06 February 2018*

***Citation:***
*Ravignani A, Thompson B and Filippi P (2018) The Evolution of Musicality: What Can Be Learned from Language Evolution Research? Front. Neurosci. 12:20. doi: 10.3389/fnins.2018.00020*

## INTRODUCTION

Language and music are typical human behaviors, absent in our closest living animal relatives. Since behavior and cognition do not fossilize, the earliest stages of language and music in our ancestors can only be reconstructed indirectly. Here, we argue for the value of taking into account the evolution of language when studying the evolution of music. There are multiple reasons for this. First, from a meta-scientific perspective, the empirical investigation of language evolution predates research on the evolution of music. Second, methodologically, the fields of music and language evolution show several commonalities (**Table 1**). For instance, in both fields, corpus-based research is complemented by laboratory-based psychological testing, electrophysiology and neuroimaging studies, and comparative experiments on animals. Third, many hypotheses concerning the origins of music also involve language, and vice-versa. Studying language and music within a common framework provides key insights and testable hypotheses in both disciplines. Here, we argue that anti-empiricist views on language have detrimental effects on understanding its origins and evolution. We identify equivalent anti-scientific tendencies in musicology. We suggest a few ways to counteract such effects that have been established for linguistics, proposing that similar approaches should be adopted in musicology.

Language is defined here as the ability to produce and understand verbal units within interactional communication acts. A major issue in musicology is finding an operational definition of music. Cross (2003, p.79) defines music as "embodying, entraining and transposably intentionalising time in sound and action." Following Honing et al. (2015), we distinguish the notions of "musicality"—a set of traits which evolve as constrained by our cognitive and biological system that shape musical behaviors across cultures - and

**TABLE 1 |** Disciplines which can contribute to understanding the evolution of musicality, and their correspondence with affine disciplines in the evolution of language.

| Domain | Dimension | Biological constraints | Addressed functions |
|---|---|---|---|
| Music | Ontogenetic development | - Organology and phoniatrics: music production<br>- Psychoacoustics and neurophysiology: music perception<br>- Studies on typologies and invariance of music structure: music production and perception | - Caregiver-infant emotional interaction<br>- Prosocial behaviors |
| | Geo-historic development | - Geomusicology: geographic distribution of musical patterns<br>- Archeology and ethnomusicology: historic development<br>- Artificial Intelligence: automatic recognition and generation of musical patterns | NA |
| | Phylogeny | - Paleoneurology, archaeoacoustic, sensory ethnography: reconstruction of prehistoric perception of music | - Pair bonding, territorial defense, group cohesion, sexual advertisement<br>- Emotional communication |
| Language | Ontogenetic development | - Psycholinguistics: speech production and comprehension, word and grammar acquisition | Communication |
| | Geo-historic development | Linguistics: phonology, morphology, syntax in natural languages | Communication, literature |
| | Phylogeny | - Evolutionary biology: mechanisms and selective pressures underlying the emergence of language | Group Identity Signaling |
| Common approaches to the study of music and language evolution | | - Experimental Psycholinguistics and Artificial Intelligence: cultural evolution, pattern generation, semantic associations, etc. | - Emotional communication<br>- Social interactions |

"music"—a socio-cultural artifact building on the biological inclination for musicality.

## LANGUAGE AND MUSIC: DIFFERENCES AND COMMONALITIES

In language, meaningless phonemes are concatenated into larger discrete units, such as morphemes and words, in accordance with phonological and morpho-syntactic rules. These units are arbitrarily linked to meanings and conceptual representations, which may be culturally transmitted (though sound-meaning mapping in language is not always arbitrary: Monaghan et al., 2012; Parise and Spence, 2012; Imai and Kita, 2014; Dingemanse et al., 2015, 2016). In contrast, while musical tones can be concatenated into phrases or melodies according to structural rules, they are not usually arbitrarily linked to external meanings. Furthermore, while musical melodies are typically made of discrete pitches at fixed interval scales and tonal centers (in case of tonal music), spoken language involves continuous pitch rise and falls (Jackendoff, 2009).

Despite these differences, language and music share several cognitive underpinnings. A number of studies have identified common cognitive mechanisms involved in the production and perception of structural relations in both instrumental music and propositional morpho-syntax (Fedorenko et al., 2009; Patel, 2010). For instance, timing principles are used in both language and music (Ravignani et al., 2017), where longer and louder units tend to be perceived as stressed, and changes in pitch modulation orient perception of boundaries between stressed and unstressed units (Cutler et al., 1997; Curtin et al., 2005). In addition, empirical evidence from brain imaging research indicates that amusic participants show deficits in fine-grained perception of pitch (Peretz and Hyde, 2003); patients fail to distinguish a question from a statement solely on the basis of changes in pitch direction (Patel et al., 2008; Liu et al., 2010). This observed difficulty in a sample of amusic patients supports the hypothesis that music and speech intonation share specific neural resources for processing pitch patterns (but see Ayotte et al., 2002). Further brain imaging studies report a considerable overlap in the brain areas involved in the perception of pitch and rhythm patterns in words and songs (Zatorre et al., 2002; Merill et al., 2012), and in sound pattern processing in melodies and linguistic phrases (Brown et al., 2006). In adults and children, musical training facilitates syllabic and pitch processing in language (Schön et al., 2004; Besson et al., 2007).

Similarly, in both music and verbal language, emotions are expressed through similar patterns of pitch, tempo and intensity (Scherer, 1995; Juslin and Laukka, 2003; Bowling et al., 2012). For instance, in both channels, happiness is expressed by fast speech rate/tempo, medium-high voice intensity/sound level, medium-high frequency energy, high fundamental frequency (F0)/pitch level, high F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks (Juslin and Laukka, 2003). Importantly, the use of voice modulation to express emotional information within interpersonal communication might have had adaptive value in the early species of our genus *Homo*, improving their ability to respond appropriately to survival opportunities (Mithen, 2005; Filippi, 2016; Frijda, 2016; Filippi et al., 2017a,b).

Finally, in addition to potentially shared cognitive foundations, music and language share the feature of being

socially learned from the behavioral outputs of other individuals. It is well-established that the process of transmission from one-generation to another is an important part of the cultural evolutionary process that shapes languages (Kirby et al., 2014). It is becoming clear that cultural evolutionary processes may play a similar role in shaping music (Cross, 2001, 2009; Reybrouck, 2013; Ravignani et al., 2016a; Fitch, 2017a; Ravignani and Verhoef, 2018). Cultural evolution has emerged as a unifying framework in the language sciences (Christiansen and Chater, 2008), linking the cognitive bases for language with the diversity of languages observable throughout the world. A similar approach in the study of the evolution music is desirable (Trehub, 2015).

## RESEARCH ON THE EVOLUTION OF MUSICALITY: PITFALLS TO AVOID

In the last decades, researchers from multiple disciplines have joined forces to unveil the origins and evolution of language. Despite patent progress in methodologies and applications (as attested by highly influential publications), a group of scientists, linguists and philosophers have strongly criticized the whole field of language evolution. This is particularly interesting because the "attack" has been led by one of the fathers of modern linguistics and cognitive science, Noam Chomsky.

A central theme of Chomsky's critiques to the empirical study of language is *mysterianism*: the idea that scientific knowledge is not always attainable (Chomsky, 2009, 2015; Piattelli-Palmarini et al., 2009). This would be due to the architecture of our minds imposing hard limits on what we can discover scientifically. In other words, Chomsky argues that the limits of human cognition hinder our capacity to unveil scientific mysteries through scientific investigation (Ravignani and Thompson, 2017). This perspective is not new, and had been already proposed in the cognitive sciences in general (McGinn, 1989).

Chomsky makes a distinction between I-language and E-language (Chomsky, 2015). I-language refers to internal linguistic representations, a universal language of thought also known as universal grammar, which, according to Chomsky, is innate. In his view, I-language has a dedicated brain area and is the only aspect of language worth studying, since it is not subject to variation through time and cultures. From his perspective, the ease with which infants acquire any of the hugely diverse set of grammatical structures observable across natural languages points to the existence of this innate universal grammar underlying language acquisition. Instead, E-language encompasses the multiple languages, i.e., the strings of sounds uttered in the outside world, which vary across individuals and cultures. Chomsky claims that processes underlying development and evolution of the wide variety of E-language instances cannot be investigated empirically. In fact, his ideas of mysterianism applied to language bring Chomsky to a simple conclusion: the most important questions about language (his I-language), including its cultural nature, its origins and development over time, and its acquisition are potentially unanswerable (Chomsky, 2015; Ravignani and Thompson, 2017). Notably, the I-language vs. E-languages distinction parallels that between *musicality*, the human cognitive-biological innate predispositions underlying music perception and production vs. *music*, intended as a cultural product (Honing et al., 2015; Honing, 2017).

Chomsky has a strong influence on debates about language evolution, and there are clear parallels between the fields of music and language evolution. Here we suggest in the strongest possible terms that research on the evolution of music should avoid Chomskyan mysterianism. By definition this perspective is scientifically stagnant, and the theoretical commitments driving Chomsky's mysterianism are widely rejected in the language sciences at large (Boeckx and Theofanopoulou, 2015; Corballis, 2017; Fitch, 2017b; Kirby, 2017; Ravignani and Thompson, 2017). Other schools of thought show some commonalities to mysterianism (McClelland, 2013; Coleman, 2016).

Another potential attack to the approach we propose comes from a field far away from Chomskyan thought. In particular, generations of ethnomusicologists and cultural anthropologists have strongly opposed the very idea of investigating evolutionary and cross-cultural features of music (Vandor, 1980; Nettl, 2005; Nattiez, 2012). These scholars often object that musical cultures from different parts of the world cannot be compared because comparison would occur through the eyes of a scholar bound to a specific culture. They argue that, across cultures, the word "music" maps to different meanings, and likewise what we call music in the Western world can be translated in many different ways across cultures and languages. According to most of these scholars, the concept of "universal"—corresponding to a feature which is found more often than not, above chance, across cultures, e.g., music often entailing percussion instruments—is pointless because it conflicts with cultural specificity. More generally, several scholars argue there is an irreconcilable divide between humanities and sciences (Gourlay, 1984; Cohen, 2001).

Both Chomskyan and anthropological schools of thought, while departing from opposite philosophical stances, reach the same theoretical conclusion: the nature of language and music is mostly unknown, and empirical efforts to unveil it are pointless. We strongly disagree with this conclusion. Even a cursory glance at the contemporary literature on language and cognition reveals astounding progress in what we know about these topics (Fitch, 2017c). Here we simply reiterate that this progress mostly results from broad contemporary adoption of the scientific method. Our view is that music-related disciplines can benefit equally from rejection of mysterianist skepticism and continued adoption of an integrated experimental approach, namely: (1) observing behaviors and the environment in which they occur, (2) formulating a hypothesis, (3) testing the hypothesis by performing an experiment or collecting new data, (4) using the results to build a model of the phenomenon of interest, (5) employing the model to generate a more refined hypothesis to be tested empirically.

Against the Chomskyan and anthropological perspectives above, we argue for an empirical approach to the origin and evolution of music and musicality. The new-born discipline of music evolution will benefit—and has already benefitted—by unifying the following approaches into one research framework: (1) empirical investigations, as opposed to armchair speculation,

on ontogenetic and phylogenetic evolution of music; (2) comparative research, addressed by probing for presence of proto-musical behaviors in other animal species; (3) cross-cultural work, recognizing the diversity of world musical behaviors while comparing them to find common patterns; (4) proposing alternatives to the classical nature-nurture dichotomy. Below we discuss these four points succinctly.

## REPLACING ARMCHAIR SPECULATION WITH MODELS

First, centuries of scientific practice have shown that tightly integrating theory and empirics typically leads to scientific progress. The study of the evolution of music or language is not an exception to this. Theoretical frameworks *should* provide testable empirical questions (Iversen, 2016), insights for good experimental design and conceptual frameworks to interpret statistical results and generate new testable hypotheses. Ideally, theoretical contributions should be formulated as mathematical models and computer simulations. One advantage of modeling, as compared to constraint-free theorizing, is modeling's potential for falsifiability: Models rely on assumptions to make predictions, which can be empirically falsified. Another related advantage of models is their potential integration with experiments: model assumptions and predictions can be promptly translated into experimental constraints, in turn testable on humans or other living organisms. For these reasons, we strongly support the use of quantitative models of cognitive (Perfors et al., 2011; Fitch, 2014), cultural (Tamariz and Kirby, 2016), and evolutionary (Thompson et al., 2016) processes in music evolution research.

## COMPARATIVE COGNITION CAN INFORM HUMAN EVOLUTION

Second, the comparative approach to animal cognition can be useful in reconstructing the evolution of human behaviors. Proto-musical behaviors may emerge in other species by (1) homology, i.e. our last common ancestor with that species was endowed with a predisposition toward the behavior under scrutiny, or (2) convergent evolution, i.e., similar evolutionary pressures gave rise to similar genetic predisposition for proto-musical behaviors in humans and other species (Fuhrmann et al., 2014; Ravignani et al., 2014, 2016b; Wilson and Cook, 2016). For instance, recent studies found evidence for beat perception and production, relative pitch and tonal encoding (Hoeschele et al., 2015; Hoeschele and Bowling, 2016), octave generalization (Crickmore, 2003), and consonance (Cook and Fujisawa, 2006) in animals. Based on theoretically driven empirical research (Honing et al., 2015), we argue that, if musical tasks designed for humans are adapted - by modifying their form, not substance—to the specific species under inquiry, many "unthinkable tasks"—sensu Chomsky (2015)—may become manageable. In contrast, based on purely theoretical introspection, Chomsky argues that the cognitive differences between humans and other animals are a matter of quality, not quantity. For instance, he claims that "rats cannot deal with a prime number maze" (Chomsky, 2015; p. 105). To this, we respond that cicadas show behavioral patterns based on prime numbers (Grant, 2005; Tanaka et al., 2009); hence it is tenable that these insects could solve a cicada-adapted prime maze, say over an evolutionary timescale. Likewise, if precursors to music are studied across species with an open-minded attitude, suggestive parallels can be found across humans and other species.

## CROSS-CULTURAL COMPARISONS

Third, an equally important sort of comparative approach consists in cross-cultural research on music. Acceptance of cross-cultural work exhibits one of the starkest contrasts between the study of language and that of music. Cross-cultural comparison of languages has proceeded unhindered for centuries, with few roadblocks. The comparative study of music, instead, experienced a golden period before a crashing halt in the 1960s. Music research can learn from evolutionary linguistics, by performing more cross-cultural work. Its purpose is to account for the uniformity and diversity of musical forms across cultures, in turn to find patterns of music and musicality which are truly generalizable to mankind (the so called universals; Savage et al., 2015).

## BEYOND THE NATURE-NURTURE DICHOTOMY

Fourth, language research has been historically characterized by an overreliance on the nature-nurture dichotomy. In our view, the evolution of language as a scientific endeavor has long been plagued by a spurious dichotomy that need not be imported into the evolution of music (Fitch, 2011). We see this opposition as reflecting unnecessarily strong but historically entrenched *theoretical* divisions. On the one hand, generative linguists in the Chomskyan school have traditionally sought an exclusively mind-internal explanation for the complex tapestry of structures and operations observable in languages (although Chomsky's contemporary mysterianism questions the future of this endeavor). On the other hand, empiricists have privileged the mechanisms of human interaction and cultural transmission as explanations for the diversity and complexity of languages. These opposing standpoints have fuelled much debate but little consensus (e.g., Carruthers et al., 2005).

Contemporary nativist-empiricist dialogues in the cognitive sciences at large now focus on developing paradigms that generalize this dichotomy into a continuous range of possibilities, with traditional nativism and empiricism at the extremes. One example of such a paradigm is Bayesian cognitive modeling, in which bio-cognitive constraints and empirical learning are integrated by a theory of subjective inference based on the principles of conditional probability. The Bayesian approach provides a framework in which both empirical evidence and innate (or earlier learned) biases can be expressed as influences on how individuals learn. While earlier approaches have treated these influences as distinct alternatives, the Bayesian approach allows the two to be balanced in a formally explicit way. Several

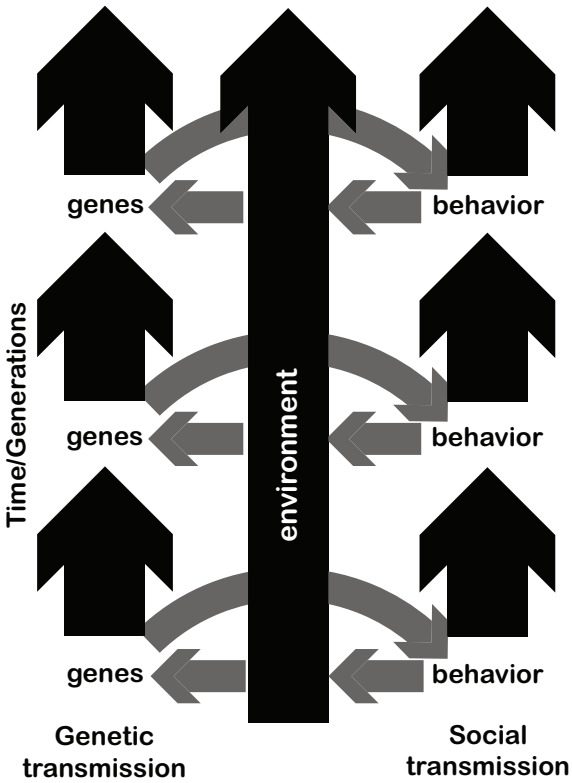



**FIGURE 1 |** The arrows depict the potential three-way interaction between genes, environment and musical behaviors. Figure inspired by Deacon (1998). Over time, both genes **(left)** and behaviors **(right)** change and are transmitted. The environment might change stochastically, or might be directionally altered by behaviors of the species populating it. For a given generation or time period, genes affect behavioral patterns. Behavioral patterns also adapt to, and modify, their environmental medium. In turn, behaviorally-driven changes in the environment might affect the fitness landscape of a species, influencing in turn which genes will be passed on to the next generation.

scholars have argued that approaches like this, in which any flavor of theory can be formally instantiated, interrogated, and tested against empirical data, are the future of nativist-empiricist dialogues (e.g., Spelke and Kinzler, 2009). We hope this sort of integrative approach can be imported into the evolution of music (Trehub, 2015; Jacoby and McDermott, 2017; Ravignani et al., accepted). For example, the generative program applied to music (Lerdahl and Jackendoff, 1985; Rohrmeier, 2011) has generated a range of predictions which have been tested empirically (e.g., Koelsch et al., 2013).

In the language sciences, a new generation of researchers is pushing this approach, integrating models and empirics, even further toward larger-scale evolutionary questions (Blasi et al., 2017). This effort focuses on process-based explanations for behavior, allowing us to understand our species a unique *interaction* of bio-cognitive and cultural processes (**Figure 1**) via evolutionary modeling (Smith and Kirby, 2008). Modeling the evolution of minds as part of cultural systems enables us to formalize theories concerning optimal divisions of labor between specialized individual minds and the cultural processes that connect them (Thompson et al., 2016; de Boer and Thompson, 2018). Free from theoretical commitments to one evolutionary process dominating another, contemporary language scientists have at their explanatory disposal an evolutionary framework more powerful than exclusively biological or cultural explanations for behavior: co-evolution, in which evolving minds shape new behaviors, and evolving behaviors shape new minds. Like the Bayesian paradigm, co-evolutionary approaches to the origins of our abilities allow us to move past the idea that only biology or only culture is relevant to unique human behaviors, in a formally explicit way. Hence, modern approaches to the study of language evolution (e.g. Kirby et al., 2014; Thompson et al., 2016) show how biological and cognitive theory can be used to develop process-based, experimentally testable hypotheses about the emergence of behavior among culturally interacting individual minds (Trehub, 2015; Ravignani et al., accepted).

## CONCLUSIONS

We hope the evolution of music can hit the ground running under by adopting the inclusive approach described above (see Bowling et al., 2017). The main hypothesis under scrutiny is: does our species unique blend of biological and cultural features underpin remarkable human behaviors like language and music (**Figure 1**)?

The distinction between music and musicality provides a practical advantage in designing and interpreting experiments. Still, empirical research on the origins of music should adopt a hybrid approach, complementing experiments in tightly controlled settings—hence targeting only music or musicality—with research which integrates the two domains and explanatory levels.

Above we only describe the importance of one possible flow of ideas, from language to music. However, the inverse is also needed: More linguists would need to learn about music evolution and cognition research. For instance, phoneticians and phonologists should capitalize on music findings when investigating tone, prosody, rhythm, etc. (Brown, 2017).

In sum, recent debates provide a bird's-eye view of how the science of language has historically developed, and partly branched into the stormy study of language evolution. Scientists addressing music can benefit from historical breakthroughs and dead-ends in the study of language evolution, and use these insights to accelerate discovery in one of the most exciting topics in contemporary cognitive science, the evolution of music.

## AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct, and intellectual contributions to the work, and approved it for publication.

## FUNDING

AR was supported by funding from the European Union's Horizon 2020 research and innovation programme under the

Marie Skłodowska-Curie grant agreement No. 665501 with the research Foundation Flanders (FWO) (Pegasus$^2$ Marie Curie fellowship 12N5517N awarded to AR) and ERC grant 283435 ABACUS (awarded to Bart de Boer). PF was supported by grants ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). All authors were supported by a visiting fellowship in Language Evolution from the Max Planck Society.

## REFERENCES

Ayotte, J., Peretz, I., and Hyde, K. (2002). Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125, 238–251. doi: 10.1093/brain/awf028

Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.* 25, 399–410.

Blasi, D. E., Michaelis, S. M., and Haspelmath, M. (2017). Grammars are robustly transmitted even during the emergence of creole languages. *Nat. Hum. Behav.* 1, 723–729. doi: 10.1038/s41562-017-0192-4

Boeckx, C., and Theofanopoulou, C. (2015). Commentary on: Labels, cognomes, and cyclic computation: an ethological perspective. *Front. Psychol.* 6:784. doi: 10.3389/fpsyg.2015.00784

Bowling, D. L., Hoeschele, M., Gill, K. Z., and Fitch, W. T. (2017). The nature and nurture of musical consonance. *Music Percept*. 35, 118–121. doi: 10.1525/mp.2017.35.1.118

Bowling, D. L., Sundararajan, J., Han, S., and Purves, D. (2012). Expression of emotion in eastern and western music mirrors vocalization. *PLoS ONE* 7:e31942. doi: 10.1371/journal.pone.0031942

Brown, S. (2017). A joint prosodic origin of language and music. *Front. Psychol.* 8:1894.

Brown, S., Martinez, M. J., and Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *Eur. J. Neurosci.* 23, 2791–2803. doi: 10.1111/j.1460-9568.2006.04785.x

Carruthers, P., Laurence, S., and Stich, S., (eds.). (2005). *The Innate Mind: Structure and Contents*, Vol. 1. New York, NY: Oxford University Press.

Chomsky, N. (2009). The mysteries of nature: how deeply hidden? *J. Philos.* 106, 167–200. doi: 10.5840/jphil2009106416

Chomsky, N. (2015). *What Kind of Creatures Are We?* New York, NY: Columbia University Press.

Christiansen, M., and Chater, N. (2008). Language as shaped by the brain. *Behav. Brain Sci.* 1, 489–508. doi: 10.1017/S0140525X08004998

Cohen, B. R. (2001). Science and humanities: across two cultures and into science studies. *Endeavour* 25, 8–12. doi: 10.1016/S0160-9327(00)01335-1

Coleman, S. (2016). "Panpsychism and neutral monism: how to make up one's mind," in *Panpsychism: Contemporary Perspectives . Philosophy of Mind Series*, eds G. Bruntrup and L. Jaskolla (New York, NY: OUP), 249–282.

Cook, N. D., and Fujisawa, T. X. (2006). The psychophysics of harmony perception: harmony is a three-tone phenomenon. *Empirical Musicol. Rev.* 1, 106–126. doi: 10.18061/1811/24080

Corballis, M. C. (2017). Language evolution: a changing perspective. *Trends Cogn. Sci.* 21, 229–236. doi: 10.1016/j.tics.2017.01.013

Crickmore, L. (2003). A re-valuation of the ancient science of harmonics. *Psychol. Music* 31, 391–403. doi: 10.1177/03057356030314004

Cross, I. (2001). Music, cognition, culture, and evolution. *Ann. N. Y. Acad. Sci.* 930, 28–42. doi: 10.1111/j.1749-6632.2001.tb05723.x

Cross, I. (2003). Music and evolution: consequences and causes. *Contemp. Music Rev*. 22, 79–89. doi: 10.1080/0749446032000150906

Cross, I. (2009). The evolutionary nature of musical meaning. *Music. Sci.* 13, 179–200. doi: 10.1177/1029864909013002091

Curtin, S., Mintz, T. H., and Christiansen, M. H. (2005). Stress changes the representational landscape: evidence from word segmentation. *Cognition* 96, 233–262. doi: 10.1016/j.cognition.2004.08.005

Cutler, A., Oahan, D., and Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Lang. Speech* 40, 141–201. doi: 10.1177/002383099704000203

Deacon, T. W. (1998). *The Symbolic Species: The Co-evolution of Language and the Brain*. New York, NY: WW Norton & Company.

de Boer, B., and Thompson, B. (2018). Biology-culture co-evolution in finite populations. *Sci. Rep.* 8:1209. doi: 10.1038/s41598-017-18928-0

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., Mitterer, H., Akita, K., et al. Woodworth, N. L. (2016). What sound symbolism can and cannot do: testing the iconicity of ideophones from five languages. *Language* 92, e117–e133. doi: 10.1353/lan.2016.0034

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., and Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Mem. Cogn.* 37, 1–9. doi: 10.3758/MC.37.1.1

Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Front. Psychol.* 7:1393. doi: 10.3389/fpsyg.2016.01393

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., et al. (2017a). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proc. R. Soc. B.* 284:20170990. doi: 10.1098/rspb.2017.0990

Filippi, P., Ocklenburg, S., Bowling, D. L., Heege, L., Güntürkün, O., Newen, A., et al. (2017b). More than words (and faces): evidence for a Stroop effect of prosody in emotion word processing. *Cogn. Emotion* 31, 879–891. doi: 10.1080/02699931.2016.1177489

Fitch, W. T. (2011). Biological versus cultural evolution: beyond a false dichotomy: comment on "modeling the cultural evolution of language" by Luc Steels. *Phys. Life Rev.* 8, 357–358. doi: 10.1016/j.plrev.2011.10.020

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys. Life Rev.* 11, 329–364. doi: 10.1016/j.plrev.2014.04.005

Fitch, W. T. (2017a). Cultural evolution: lab-cultured musical universals. *Nat. Hum. Behav*. 1, 1–2. doi: 10.1038/s41562-016-0018

Fitch, W. T. (2017b). On externalization and cognitive continuity in language evolution. *Mind Lang.* 32, 597–606. doi: 10.1111/mila.12162

Fitch, W. T. (2017c). Empirical approaches to the study of language evolution. *Psychon. Bull. Rev.* 24, 3–33. doi: 10.3758/s13423-017-1236-5

Frijda, N. H. (2016). The evolutionary emergence of what we call "emotions". *Cogn. Emot.* 30, 609–620. doi: 10.1080/02699931.2016.1145106

Fuhrmann, D., Ravignani, A., Marshall-Pescini, S., and Whiten, A. (2014). Synchrony and motor mimicking in chimpanzee observational learning. *Sci. Rep*. 4:5283. doi: 10.1038/srep05283

Gourlay, K. A. (1984). The non-universality of music and the universality. *World Music* 26, 25–39.

Grant, P. R. (2005). The priming of periodical cicada life cycles. *Trends Ecol. Evol.* 20, 169–174. doi: 10.1016/j.tree.2005.01.016

Hoeschele, M., and Bowling, D. L. (2016). Sex differences in rhythmic preferences in the budgerigar (*Melopsittacus undulatus*): a comparative study with humans. *Front. Psychol.* 7:543. doi: 10.3389/fpsyg.2016.01543

Hoeschele, M., Merchant, H., Kikuchi, Y., Hattori, Y., and ten Cate, C. (2015). Searching for the origins of musicality across species. *Phil. Trans. R. Soc. B* 370:20140094. doi: 10.1098/rstb.2014.0094

Honing, H. (2017). "Musicality as an upbeat to music: introduction and research agenda," in *The Origins of Musicality,* ed H. Honing (Cambridge, MA: MIT Press), 3–20.

Honing, H., ten Cate, C., Peretz, I., and Trehub, S. E. (2015). Without it no music: cognition, biology, and evolution of musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 370:20140088. doi: 10.1098/rstb.2014.0088

Imai, M., and Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130298. doi: 10.1098/rstb.2013.0298

Iversen, J. R. (2016). "In the beginning was the beat: evolutionary origins of musical rhythm in humans," in *The Cambridge Companion to Percussion*, ed R. Hartenberger (Cambridge: Cambridge University Press), 281–295.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Percept.* 26, 195–204. doi: 10.1525/mp.2009.26.3.195

Jacoby, N., and McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr. Biol.* 27, 359–370. doi: 10.1016/j.cub.2016.12.031

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychon. Bull. Rev.* 24, 118–137. doi: 10.3758/s13423-016-1166-7

Kirby, S., Griffiths, T., and Smith, K. (2014). Iterated learning and the evolution of language Introduction: can culture explain structure? *Curr. Opin. Neurobiol.* 28, 108–114. doi: 10.1016/j.conb.2014.07.014

Koelsch, S., Rohrmeier, M., Torrecuso, R., and Jentschke, S. (2013). Processing of hierarchical syntactic structure in music. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15443–15448. doi: 10.1073/pnas.1300272110

Lerdahl, F., and Jackendoff, R. (1985). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Liu, F., Patel, A. D., Fourcin, A., and Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain* 133, 1682–1693. doi: 10.1093/brain/awq089

McClelland, T. (2013). The Neo-Russellian ignorance hypothesis. *J. Conscious. Stud.* 20, 4–125. Available online at: http://www.ingentaconnect.com/content/imp/jcs/2013/00000020/F0020003/art00006?crawler=true

McGinn, C. (1989). Can we solve the mind - body problem? *Mind* xcviii 391, 349–366. doi: 10.1093/mind/XCVIII.391.349

Merill, J., Sammler, D., Bangert, M., Goldhahn, D., Turner, R., and Friederici, A. D. (2012). Perception of words and pitch patterns in song and speech. *Front. Psychol.* 3:76. doi: 10.3389/fpsyg.2012.00076

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. Cambridge, MA: Harvard University Press.

Monaghan, P., Mattock, K., and Walker, P. (2012). The role of sound symbolism in language learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1152–1164. doi: 10.1037/a0027747

Nattiez, J.-J. (2012). Is the search for universals incompatible with the study of cultural specificity? *Hum. Soc. Stud.* 1, 67–94. doi: 10.2478/v10317-012-0005-2

Nettl, B. (2005). *The Study of Ethnomusicology: Thirty-One Issues and Concepts*. Champaign, IL: University of Illinois Press.

Parise, C. V., and Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Exp. Brain Res.* 220, 319–333. doi: 10.1007/s00221-012-3140-6

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford: Oxford University Press.

Patel, A. D., Wong, M., Foxton, J., Lochy, A., and Peretz, I. (2008). Speech intonation perception deficits in musical tone deafness (congenital amusia). *Music Percept.* 25, 357–368. doi: 10.1525/mp.2008.25.4.357

Peretz, I., and Hyde, K. L. (2003). What is specific to music processing? insights from congenital amusia. *Trends Cogn. Sci.* 7, 362–367. doi: 10.1016/S1364-6613(03)00150-5

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., and Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition* 120, 302–321. doi: 10.1016/j.cognition.2010.11.015

Piattelli-Palmarini, M., Uriagereka, J., and Salaburu, P. (Eds.). (2009). *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. OUP Oxford.

Ravignani, A., and Thompson, B. (2017). NOAM CHOMSKY, What kind of creatures are we? New York, NY: Columbia University Press, 2015. Pp. 200. Hb. $19.95. *Lang. Soc.* 46, 446–447. doi: 10.1017/S0047404517000288

Ravignani, A., and Verhoef, T. (2018). Which melodic universals emerge from repeated signaling games? *Artif. Life* 24. doi: 10.7287/peerj.preprints.3366v1

Ravignani, A., Bowling, D. L., and Fitch, W. T. (2014). Chorusing, synchrony and the evolutionary functions of rhythm. *Front. Psychol.* 5:1118. doi: 10.3389/fpsyg.2014.01118

Ravignani, A., Delgado, T., and Kirby, S. (2016a). Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* 1:0007. doi: 10.1038/s41562-016-0007

Ravignani, A., Fitch, W. T., Hanke, F. D., Heinrich, T., Hurgitsch, B., Kotz, S. A., et al. (2016b). What pinnipeds have to say about human speech, music, and the evolution of rhythm. *Front. Neurosci.* 10:274. doi: 10.3389/fnins.2016.00274

Ravignani, A., Honing, H., and Kotz, S. A. (2017). The evolution of rhythm cognition: Timing in music and speech. *Front. Hum. Neurosci.* 11:308. doi: 10.3389/fnhum.2017.00303

Ravignani, A., Thompson, B., Grossi, T., Delgado, T., and Kirby, S. (accepted). Evolving building blocks of rhythm: How human cognition creates music via cultural transmission. *Ann. N.Y. Acad. Sci.* doi: 10.1111/nyas.13610

Reybrouck, M. (2013). From sound to music: an evolutionary approach to musical semantics. *Biosemiotics* 6, 585–606. doi: 10.1007/s12304-013-9192-6

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *J. Math. Music* 5, 35–53. doi: 10.1080/17459737.2011.573676

Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8987–8992. doi: 10.1073/pnas.1414495112

Scherer, K. R. (1995). Expression of emotion in voice and music. *J. Voice* 9, 235–248. doi: 10.1016/S0892-1997(05)80231-0

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349. doi: 10.1111/1469-8986.00172.x

Smith, K., and Kirby, S. (2008). Cultural evolution: implications for understanding the human language faculty and its evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3591–3603. doi: 10.1098/rstb.2008.0145

Spelke, E. S., and Kinzler, K. D. (2009). Innateness, learning, and rationality. *Child Dev. Perspect.* 3, 96–98. doi: 10.1111/j.1750-8606.2009.00085.x

Tamariz, M., and Kirby, S. (2016). The cultural evolution of language. *Curr. Opin. Psychol.* 8, 37–43. doi: 10.1016/j.copsyc.2015.09.003

Tanaka, Y., Yoshimura, J., Simon, C., Cooley, J. R., and Tainaka, K.-I. (2009). Allele effect in the selection for prime-numbered cycles in periodical cicadas *Proc. Natl. Acad. Sci. U.S.A.* 106, 8975–8979. doi: 10.1073/pnas.0900215106

Thompson, B., Kirby, S., and Smith, K. (2016). Culture shapes the evolution of cognition. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4530–4535. doi: 10.1073/pnas.1523631113

Trehub, S. E. (2015). Cross-cultural convergence of musical features. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8809–8810. doi: 10.1073/pnas.1510724112

Vandor, I. (1980). Note de la rédaction. *World Music* 1:3.

Wilson, M., and Cook, P. F. (2016). Rhythmic entrainment: why humans want to, fireflies can't help it, pet birds try, and sea lions have to be bribed. *Psychon. Bull. Rev.* 23, 1647. doi: 10.3758/s13423-016-1013-x

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* 6, 37–46. doi: 10.1016/S1364-6613(00)01816-7

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2018 Ravignani, Thompson and Filippi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

GENERAL COMMENTARY
published: 26 February 2018
doi: 10.3389/fpsyg.2018.00075

# Commentary: The 'Musilanguage' Model of Language Evolution

*Aleksey Nikolsky**

*Braavo Enterprises, Los Angeles, CA, United States*

**Keywords: texture, heterophony, polyphony, homophony, musilanguage, asynchrony, isophony, meter**

**A commentary on**

**The 'Musilanguage' Model of Language Evolution**
*by Brown, S. (2000). The Origins of Music. eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 271–300. doi: 10.1037/e533412004-001*

The model of musilanguage (Brown, 2000, 2017) requires a new musicological term to refer to its texture. Like choral singing, and unlike speech, musilanguage is based on simultaneous vocalization of multiple participants who reproduce the same signal (call) at random time intervals and pitch levels, akin to a wolf "chorus.[1]" Voiced utterances produce multiple pitches, generating a "jumbled" texture, similar to polyphony and heterophony, but not fully qualifying as such.

The term "heterophony" was introduced by Stumpf (1897) to refer to the fusion of sounds whose components retained their singular identity. Stumpf discovered this word in Plato's *Laws* (Plato, 2013, p. 203), where it referred to pitch and rhythm discrepancy between the vocals and the lyre performing the same tune. Four years later, Stumpf (1901) reused this term to describe Thai music. He characterized *heterophony* as the looping of simultaneous melodic paraphrases [Umspielen], where parts generally followed the same melodic contour while differing in detail, so that minute discrepancies would meet again in unison.

Adler (1908) conceptualized heterophony as a style alternative to homophony and polyphony, applicable across Siamese, Japanese, Javanese, and Russian musics. He specifically found Russian heterophony to present a paradigm of heterophonic arrangement, designed to make melodic repetition less monotonous and more idiosyncratic for each singer's voice.

The Russian Musical Encyclopedia defines heterophony as a multi-part music generated by the collective performance of the same melody, where parts contain deviations from the principal melodic formula (Mueller, 1973). Such organization is regarded as a general textural type of ornamental, harmonic, and/or polyphonic variation that can complicate classification. Indeed, already in 1911, Stumpf criticized Adler for misapplication of the term (Stumpf and David, 2012). The keynote of heterophony is an ongoing melodic repetition with numerous intermittent variations, which seems to apply to musilanguage chorusing (Brown, 2007). However, such chorusing contains no synchronization, whereas heterophony implies prevalent synchronization of parts.

Swan (1943) defined heterophony as "a *principal melody* improvised *simultaneously* by several singers, *retaining its main outline in each voice*, yet showing enough independence to result *in places* in 2- and 3- and even 4-part harmony[2]." The Grove Dictionary (Cooke, 2001), following Swan's definition, emphasizes a collective synchronized execution. Although the notation example provided in the Grove article shows a consistent misalignment of 5 parts

[1]The sample of a wolf pack howling can be heard at:http://chirb.it/kFLxIC. It is characterized by the reproduction of the same call at various speeds and pitch levels, without any synchronization of the onset and termination points of the calls.
[2]I have italicized those words in the definition that imply the general coordination in time between the performances of all the participants.

OPEN ACCESS

***Edited by:***
*Timothy L. Hubbard,*
*Arizona State University, United States*

***Reviewed by:***
*Steven Brown,*
*McMaster University, Canada*

***Correspondence:**
*Aleksey Nikolsky*
*aleksey@braavo.org*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *09 November 2017*
***Accepted:*** *18 January 2018*
***Published:*** *26 February 2018*

***Citation:***
*Nikolsky A (2018) Commentary: The 'Musilanguage' Model of Language Evolution. Front. Psychol. 9:75. doi: 10.3389/fpsyg.2018.00075*

**TABLE 1 |** Feature comparison of 4 types of musical texture in a multi-part ensemble.

| Parts | Heterophony | Polyphony | Homophony | Isophony |
|---|---|---|---|---|
| Numerosity | From 2 to a few dozen parts—basically, as many as there are participants. Each part is performed solo, but often merges with other parts to make a tutti unison. | From 2 to 3 (or 4) parts. In Western art-music, usually up to 8 (and rarely up to 40). A part can be performed by a soloist or a group. Each part is sustained throughout the entire music. | Usually from 2 to 5 parts—to make harmonic intervals or chords: chunks of tones, perceived as a single sonic entity. A part can be performed by a soloist or a group. It can cede and/or reappear later in music. | From 2 to an unlimited number of parts—as many as there are participants. Each part is performed solo, and it neither merges nor concords with other parts. |
| Thematicity | All parts carry out the same melody (tune)—yet, each with its own melodic details. The entire music sounds as numerous repetitions of the same theme. | Parts can use a single melody (imitational polyphony) or multiple melodies (contrasting polyphony), both of which sound continuously changing. | One part carries out the melody against the other parts that form a uniform accompaniment. The melody usually contains diverse thematic material, but can be repetitive. | All parts reproduce the same call (phrase) with variations in timing and pitch. The overall sound is more cluttered (less repetitive) than in heterophony. |
| Synchrony | Parts are synced at the initiation and termination points of the melodic phrases. | Parts are synced all the way from the beginning to the end of the music, in the order of a beat. | Parts are synced to make chords and harmonic intervals—with precision of 30–100 milliseconds. | Parts are not synced at all—each part freely proceeds at its own pace. |
| Pitch relation affordances | Each part can comprise any harmonic interval in relation to another part—but stays mostly in unison. | Each part is restricted to specific harmonic intervals in relation to another part—mostly non-unison. | Each part is restricted to specific harmonic intervals in relation to another part—mostly 5th, 3rd, 6th, and more seldom–4th. | Each part comprises any interval in relation to another part—tuned and untuned. |
| Time relation affordances | Parts stay metrically coordinated, while allowing each part to have its own rhythmic variation and expressive timing (e.g., rubato, ritenuto, fermata). | Parts stay metrically and rhythmically coordinated by means of the metric division system, leaving very limited margins for expressive timing per part. | Parts stay metrically and rhythmically coordinated, while letting melody and accompaniment each have their own autonomous rhythmic variation and expressive timing. | Parts stay metrically and rhythmically uncoordinated, each with its own variation and expressive timing. |
| Intentionality | The relationship between parts is not determined by the performers, but is rather, an emergent property of their attempt to perform the same melody. | The relationship between parts is governed by a complex set of melodic and harmonic rules that constrain the performers in their performance. | The relationship between the accompanying parts and the melodic part is governed by a simpler set of harmonic rules, binding the performers. | The relationship between parts is not consciously controlled by any of the participating performers, but rather emerges purely per chance. |
| Functionality | All parts are functionally identical in presenting the same melodic idea, with minute variations in pitch and rhythm. Each part is equal in status to other parts, where they all collaborate to continuously maintain the same expression throughout the music in a rather static way. | Imitative polyphony presents a single melodic idea expressed in succession of "call" and "response," where "call" subdues "response." Contrasting polyphony presents multiple melodies with diverse functions. All polyphonic parts compete for attention, promoting melodic development. | Melodic part presents the principal idea which is usually followed by secondary melodic ideas. Other parts can all carry the same accompanying function, or be specialized (e.g., bass, pedal, melodic figuration). Parts can switch their functions and regroup—thereby generating contrast, development, and recapitulation. | All parts contain the same call that is repeated with substantial changes in frequency and duration. Each part is equal in status to other parts in producing an independent expression of the same idea by each performer—in the manner of reaffirming this idea. |

(Knudsen, 1968), such heterophony is rare and does not sound "jumbled[3]." Its asynchronicities remain minimal (<half-a-beat). Longer asynchronies (≥beat) generate polyphonic imitations, where the same melody becomes deliberately distributed between multiple parts to produce juxtapositions at certain temporal intervals[4].

*Polyphony* is "a style of simultaneously combining number of parts, each forming an individual melody and harmonizing with each other" (Oxford Dictionary). Despite its association with Western art-music, polyphony penetrated Western popular (Bukofzer, 1940) and traditional music (Ahmedaja, 2011), prompting research of non-European polyphony (Arom, 2004; Jordania, 2006). Many ethnomusicologists prefer alternative terms (diaphony, disphony), while others treat "polyphony" as an umbrella term for any multi-part music, setting terminological confusion (Cooke, 2001). Current consensus defines polyphony as "a mode of expression based on simultaneous combination of *separate parts*, perceived and produced *intentionally* in their *mutual differentiation*, in a given *formal order*" (Agamennone, 1996)[5].

[3]An example of the Hebrides Psalm can be heard at: http://chirb.it/raeaOr. It presents a clear single melodic line that is "smudged" by consistent delays between different parts—presenting a rare style of melodic variation by means of a deliberate "reverberation" effect induced by the continuous sub-beat time lagging between parts.

[4]Well-known examples of musical texture in monothematic imitational polyphony in folk music are round-songs. In art-music, Bach's "Musical Offering" BWV 1079 contains excellent demonstration of polyphonic arrangement of a single melody in the so-called "riddle canons"—notated as a melody solo ought to be performed by a few musicians who enter one by one in specific time intervals.

[5]My italics emphasize points that distinguish polyphonic texture from heterophonic.

Polyphonic and heterophonic textures differ in orderliness: heterophonic parts are inadvertent, unlike polyphonic parts (Tallmadge, 1984). Polyphony induces individualization of parts by means of sharpening their functional contrast in texture. Hence, synchronization is even more important for polyphony—parts must align in pitch and time throughout the entirety of music. This makes polyphonic performance metrically stricter than heterophonic performance.

Even stricter is synchronization in another "classic" texture—*homophony*—"music in which all melodic parts move together at more or less the same pace" (Hyer, 2001). Contrary to common belief, homophony is *not* bound to European music alone (Nikolsky, 2016). Its reliance on chords and harmonic intervals demands high concision in tones' onsets: in the order of under 100 msec (Huron, 2001), typically, 30–50 msec (Rasch, 1988).

All "classic" textures rely on *harmonic, metric, and thematic integrity of parts*. Performers attune their performance to the pitch of their partners, the manifestation of beat in their rhythms, and the distribution of musical material across parts—what musicologists call "thematic material" and consider an expressive point of a musical work by which it can be remembered (Drabkin, 2001). In this semiotic sense (Réti, 1951), the notion of *thematicity* is applicable to folk and non-Western music (Mazel, 1960). However, harmonicity, metricity, and thematicity are inadmissible for musilanguage. Even modifying "classic" terms (e.g., "jumbled heterophony") would constitute a misnomer: musilanguage inherently lacks any form of arrangement of parts.

Since musilanguage occupies an evolutionary position between the "natural" animal vocalizations and the simplest human oral communication, it predates mode, scale, meter—and therefore, heterophony and polyphony. This situation calls for a new term—*isophony:* texture that uses brief calls, continuously reproduced by multiple performers with irrational deviations in timing and pitch, where each participant retains idiosyncrasy of the rhythmic, timbral, and directional attributes of the pitch contour—altogether producing a "jumbled" effect (Nikolsky, 2016, Appendix-5)[6]. Vocalization can be considered "isophonic" if it maintains a single call as a unit of texture, scalable shorter/longer and higher/lower through the continuum of duration and frequency for every participating part—consistently reproducing that call out-of-sync in relation to the moment of its onset or termination.

Isophony contrasts "classic" textures by its tendency to expose each participant's identity without enmeshing into the ensemble. Isophony involves the *assembly of individuals*, rather than a single entity ("choir"). Isophonic tones never meet in unison or in beat, and are devoid of any form of harmonization[7]. Isophony's only feature of tonal organization is the uniformity of the melodic and timbral characteristics of a call. The function of isophonic texture is to attract attention to each participant's expression of the same state of mind. The important features that distinguish isophony from heterophony, polyphony, and homophony are summarized in the **Table 1**.

Conceptualization of isophony as a primordial texture that predated music, establishes the lineage in the morphological evolution of music, allowing comparative cross-examination of musical structures in multi-part music.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

[6]This appendix contains an overview of the evolution of musical texture, following the model of "histories of fine art"—see Table 2 (p. 13–16) in it.

[7]An example of isophony is an Akia tribe song (courtesy of Anthony Seeger): http://chirb.it/ryANbJ. The leader coins the call (vocable "Tete"), and the tribe repeats it at various pitch levels and duration values, where each participant displays their specific social status through the relative duration of the same call (Seeger, 2004).

## REFERENCES


Adler, G. (1908). "Über heterophonie," in *Jahrbuch Der Musikbibliothek Peters XV*, ed R. Schwartz (Leipzig: Edition Peters), 17–27.

Agamennone, M. (1996). *Polifonie. Procedimenti, Tassonomie E Forme: Una Riflessione "a Più Voci."* Venice: Edizioni Il Cardo.

Ahmedaja, A. (2011). *European Voices: Cultural Listening and Local Discourse in Multipart Singing Traditions in Europe, Vol. 2.* Vienna; Koln; Weimar: Böhlau Verlag.

Arom, S. (2004). *African Polyphony and Polyrhythm: Musical Structure and Methodology.* Transl. by M. Thom, B. Tuckett, and R. Boyd. Cambridge: Cambridge University Press.

Brown, S. (2000). "The 'Musilanguage' model of language evolution," in *The Origins of Music,* eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 271–300.

Brown, S. (2007). Contagious heterophony: a new theory about the origins of music. *Music. Sci.* 11, 3–26. doi: 10.1177/102986490701100101

Brown, S. (2017). A joint prosodic origin of language and music. *Front. Psychol.* 8:1894. doi: 10.3389/fpsyg.2017.01894

Bukofzer, M. F. (1940). Popular polyphony in the middle ages. *Music. Quart.* 26, 31–49. doi: 10.1093/mq/XXVI.1.31

Cooke, P. (2001). "Polyphony. Non-Western. General," in *The New Grove Dictionary of Music and Musicians*, eds S. Sadie and J. Tyrrell (London: Macmillan Publishers).

Drabkin, W. (2001). "Theme," in *New Grove Dict. Music Music,* eds S. Sadie S and J. Tyrrell (London: Macmillan Publishers).

Huron, D. (2001). Tone and voice: a derivation of the rules of voice-leading from perceptual principles. *Music Percept.* 19, 1–64. doi: 10.1525/mp.2001.19.1.1

Hyer, B. (2001). "Homophony," in *New Grove Dict. Music Music*, eds S. Sadie and J. Tyrrell (London: Macmillan Publishers).

Jordania, J. (2006). *Who Asked the First Question? The Origins of Human Choral Singing, Intelligence, Language and Speech.* Tbilisi: Logos.

Knudsen, T. (1968). Ornamental hymn/psalm singing in Denmark, the Faroe Islands and the hebrides. *DFS Inform.* 2, 5–22.

Mazel, L. (1960). *Structuring of the Music Works [Строение МузыкаЛьных Произведений], 2nd Edn.* Moscow: Muzyka.

Mueller, T. F. (1973). "Heterophony," in *Encyclopedia of Music, Vol. 1*, ed Y. Keldysh (Moscow: Soviet Encyclopedia), 1973–1976.

Nikolsky, A. (2016). Evolution of tonal organization in music optimizes neural mechanisms in symbolic encoding of perceptual reality. Part-2: Ancient to seventeenth century. *Front. Psychol.* 7:211. doi: 10.3389/fpsyg.2016.00211

Plato (2013). *Laws*. New York, NY: Dover Publications.

Rasch, R. A. (1988). "Timing and synchronization in ensemble performance," in *Generative Processes in Music; The Psychology of Performance, Improvisation, and Composition*, ed J. A. Sloboda (Oxford, UK: Clarendon Press), 70–90. Available online at: http://psycnet.apa.org/record/1988-98076-004

Réti, R. (1951). *The Thematic Process in Music*. London; New York, NY: Macmillan Publishers.

Seeger, A. (2004). *Why Suyá Sing: A Musical Anthropology of an Amazonian People*. Chicago, IL: University of Illinois Press.

Stumpf, C. (1897). *Die Pseudo-Aristotelischen Probleme Über Musik. Abhandlungen Der Königlichen Akademie Der Wissenschaften Zu Berlin, Vol. 3*. Berlin: Akademie der Wissenschaften.

Stumpf, C. (1901). "Tonsystem und musik der siamesen," in *Beiträge Zur Akustik Und Musikwissenschaft*, ed C. Stumpf (Leipzig: Veerlag von Johann Ambrosius Barth), 69–138.

Stumpf, C., and David, T. (2012). *The Origins of Music*. Oxford: Oxford University Press.

Swan, A. J. (1943). The nature of the Russian folk-song. *Music. Quart.* 29, 498–516. doi: 10.1093/mq/XXIX.4.498

Tallmadge, W. H. (1984). Folk organum: a study of origins. *Am. Music* 2, 47–65.

**Conflict of Interest Statement:** AN was employed by company, Braavo Enterprises as a technical and creative director. Braavo Enterprises specializes in developing content for educational and edutainment programs for children of 5–11 years of age.

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2018 Nikolsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 04 April 2018
doi: 10.3389/fpsyg.2018.00344

# "The Two Brothers": Reconciling Perceptual-Cognitive and Statistical Models of Musical Evolution

*Steven Jan**

*Music and Drama, University of Huddersfield, Huddersfield, United Kingdom*

While the "units, events and dynamics" of memetic evolution have been abstractly theorized (Lynch, 1998), they have not been applied systematically to real corpora in music. Some researchers, convinced of the validity of cultural evolution in more than the metaphorical sense adopted by much musicology, but perhaps skeptical of some or all of the claims of memetics, have attempted statistically based corpus-analysis techniques of music drawn from molecular biology, and these have offered strong evidence in favor of system-level change over time (Savage, 2017). This article argues that such statistical approaches, while illuminating, ignore the psychological realities of music-information grouping, the transmission of such groups with varying degrees of fidelity, their selection according to relative perceptual-cognitive salience, and the power of this Darwinian process to drive the systemic changes (such as the development over time of systems of tonal organization in music) that statistical methodologies measure. It asserts that a synthesis between such statistical approaches to the study of music-cultural change and the theory of memetics as applied to music (Jan, 2007), in particular the latter's perceptual-cognitive elements, would harness the strengths of each approach and deepen understanding of cultural evolution in music.

**Keywords: qualitative, quantitative, perceptual-cognitive, statistical, memetics, phylomemetics, cultural evolution**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky,*
*Independent Researcher,*
*Los Angeles, CA, United States*

***Reviewed by:***
*Piotr Podlipniak,*
*Adam Mickiewicz University in Poznań, Poland*
*Stephan Thomas Vitas,*
*Independent Researcher, Washington, DC, United States*
*Ollie Bown,*
*University of New South Wales, Australia*

***Correspondence:**
*Steven Jan*
*s.b.jan@hud.ac.uk*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *05 December 2017*
***Accepted:*** *28 February 2018*
***Published:*** *04 April 2018*

***Citation:***
*Jan S (2018) "The Two Brothers": Reconciling Perceptual-Cognitive and Statistical Models of Musical Evolution. Front. Psychol. 9:344. doi: 10.3389/fpsyg.2018.00344*

## 1. INTRODUCTION: APPROACHES TO THE STUDY OF CULTURAL EVOLUTION

The dichotomy, even tension, between qualitative and quantitative research methods aligns to some extent with the "two cultures"—the artistic/humanistic and the scientific, respectively—famously outlined by Snow (1964)[1]. While this is certainly an oversimplification—the two approaches often blend; and both may be deployed in the service of falsifiability (Popper, 1959), the acid test of a scientific theory—it has, until quite recently, largely been the norm, certainly in western musicology[2]. Nevertheless, the explosive growth in computer power, and its increasing accessibility, has, over the last two decades, put systematic approaches in the hands of scholars in the arts and humanities. In music research, such approaches are typified by the interest in "empirical

[1]I am grateful to Valerio Velardo, Alexey Nikolsky, and the anonymous reviewers for their helpful comments on earlier versions of this article.

[2]In eastern Europe, there has arguably been a more thoroughgoing synthesis, the integration of artistic/humanistic and scientific methodologies being long established, for example, in Russian scholarship (Grigoryan, 2011).

[experimental, data-rich] musicology" (Cook, 2004) and, more broadly, by the current attention paid in the humanities to the promises of "big data" (Sharma et al., 2014), which allows, for instance, for large-scale statistical analysis of music-related bibliographical data (Rose et al., 2015).

Conversely, a number of research traditions in the sciences have used music data in quantitative studies, including the Music Information Retrieval Exchange (MIREX) project[3]. This work stems partly from an interest in how technology can expedite music research—particularly in the fields of pattern-finding and data-retrieval—and partly from a recognition that the inherent complexity of music makes it a singular challenge for the design and implementation of computerized analytical tools. A similar motivation underpins cognitive science in music: often, its music-orientated practitioners pursue it in order to try to unravel the mysteries of the art form; whereas its science-orientated researchers wish to understand the deep embeddedness of music in multiple brain and body systems (Schulkin, 2013). Linking data-searching and analysis and cognitive science, the recent development of systems which autonomously create music—what might be termed the computer simulation of musical creativity—is testament to the power of computers to bring together research in music, artificial intelligence and cognitive science in the service of understanding what still seem to be the mysteries of creativity (Miranda et al., 2003; Boden, 2004), whether this research is motivated by artistic/humanistic or by scientific impulses[4].

The study of cultural evolution has been approached from both of Snow's perspectives. From the scientific, there is a tradition of research at the interface of anthropology, sociology and evolutionary biology which uses broadly Darwinian methods to understand the spread of cultural items, including ideas, artistic traditions and artifact-manufacturing technologies (Cavalli-Sforza and Feldman, 1981), this cultural transmission sometimes being correlated with genetic transmission (Shennan, 2002). From the artistic/humanistic, there is a long tradition of research (conducted broadly under the rubric of historical musicology) of referring to change in music as in some sense evolutionary (Perry, 2000). But this ascription is largely metaphorical; that is, it documents morpho-stylistic changes—in the outputs of composers, in the development of genres, or in the cultures of places or times—but it does not argue for a Darwinian (or any other algorithmic) basis as the mechanism driving this change. As an artistic/humanistic field, it clearly does not want to deny the agency of the composer, however that is understood to arise (Blackmore, 2010), just as composers do not want to deny it of and for themselves[5].

By contrast, many would argue that because musical patterns, however defined, manifestly demonstrate variation, inheritance (transmission) and selection—"principles [which] apply equally to biological and cultural evolution" (Savage, 2017, p. 9)—they conform to Darwin's theory of evolution by (natural) selection[6]. That is, such patterns—Dawkins' *memes*—instantiate the evolutionary algorithm, but are sequences of elements in cultural media—such "phemotypic" (extra-somatic) products as "tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches" (Dawkins, 1989, p. 192), which devolve to "memotypic" patterns of neuronal interconnection (Calvin, 1996; Jan, 2011; Mhatre et al., 2012)—rather than biological-medium (DNA) sequences[7]. In this sense, "music literally evolves . . . [because] musical evolution follows patterns and processes that are similar, but not identical, to [those of] genetic evolution" (Savage, 2017, pp. 38, 22).

Accepting the memetic formalization of cultural evolution as real and not metaphorical, and using a small case study which, it is hoped, can be scaled and generalized, this article attempts to reconcile approaches drawn from the perceptual-cognitive and the statistical domains as they apply to the evolution of music[8]. It regards these two domains as broadly aligning, respectively, with the qualitative/quantitative distinction discussed above, although it recognizes that the perceptual-cognitive is of course formalizable and measurable (and thus partly quantitative/statistical) using the methodologies of cognitive science. In this sense, the article emphasizes the perceptual-cognitive/statistical dichotomy as arguably more meaningful for the understanding and advancement of memetics than the qualitative/quantitative.

Section 2 discusses some of the criticisms of memetics, arguing in its defense that its central claims, grounded as they are in important psychological principles, cannot be lightly dismissed. Section 3 discusses how relationships between musical patterns can be formalized using a combination of perceptual-cognitive and statistical approaches in ways that offer a robust model for the development of memetics. Section 4 follows up some implications of memetic similarity measurements, considering the representation of evolutionary relationships using taxonomic trees. Section 5 looks forward to the future integration of perceptual-cognitive and statistical approaches using computer technology.

The article offers two principal claims. The first of these is as follows: a purely statistical approach based on counting note-edits without consideration of perceptual-cognitive aspects gives an incomplete account of cultural evolution. A second, derived, claim will be outlined at the start of section 3.

[3]http://www.music-ir.org/mirex/wiki/MIREX_HOME

[4]See also the *Journal of Creative Music Systems* (http://jcms.org.uk/).

[5]It should be remembered that the (musicological) conception of music as a series of discrete structures/objects (works, Goehr, 1992) produced by named and celebrated author-composers and notated unambiguously is a relatively recent western-European phenomenon, and that most human music is (from an *ethno*musicological standpoint) communal, processive and deeply enmeshed with other media, such as dance and poetry, and with worship. In Taruskin's (1995) phrase, this is the distinction between music as text and music as act.

[6]Because the evolutionary algorithm (Dennett, 1995, p. 343) is substrate-neutral, it makes little sense to distinguish between "natural" and "cultural" selection—this being the principle underpinning Universal Darwinism (Dawkins, 1983b).

[7]Thus, and at the risk of multiplying terminology, I use "phemotype" as the memetic counterpart to the genetic "phenotype". By extension, I use "memotype" as the counterpart to "genotype" (Jan, 2007, p. 30, Table 2.1).

[8]Thus, it takes as its starting point the assumptions that (i) culture evolves; (ii) that this evolution is broadly Darwinian; and (iii) that memetics offers the best formalization of this cultural-evolutionary process.

## 2. THE PROBLEM WITH MEMETICS?

To demonstrate the nature-culture similarities he hypothesizes, Savage (2017) uses techniques drawn from molecular genetics—discussed more fully in section 3—to compare the basic mutational-editing operations of note conservation, substitution, insertion and deletion (Savage, 2017, p. 53) in corpora of folk-song melodies with protein modification in biological transmission. He argues that an advantage of a "rigorously quantitative approach modeled on molecular genetics is that such quantitative approaches have shown success in rehabilitating cultural-evolutionary theory after much criticism of earlier incarnations such as Dawkins' "memetics"" (Savage, 2017, p. 45).

Criticism of memetics—Gould called it a "meaningless metaphor" (in Blackmore, 1999, p. 17; see also Kuper, 2000)—has arguably been counterbalanced by as much endorsement (Dennett, 2007), or at least by the acceptance that some problems in cultural studies are readily addressed by recourse to memetics. Yet Savage is to some extent correct in his implication that a fault with memetics (assuming one accepts its fundamental premises) is that it has hitherto been formulated in a somewhat imbalanced way, with too much emphasis on the qualitative and too little on the quantitative (but see McNamara, 2011). In the terms of section 1, it might therefore be believed that it has not (yet) been formulated in such a way as to be falsifiable. Yet this is to ignore the work of several scholars who have attempted to use the insights of memetics in quantitative studies (Adamic et al., 2014); and also, perhaps more importantly, to discount the work of Lynch (1998), who has arguably made the greatest contribution to the formalization of memetics, even though his models, to my knowledge, have not yet been systematically applied or tested[9].

If Savage's (2017) criticism of memetics as insufficiently orientated toward quantitative methodologies is accepted, then it is surely valuable that the more qualitative insights of memetics—often based upon introspective evaluation of the nature of certain musical patterns and their transmission across cultural time and space—are supported by quantitative work which counts and measures such phenomena systematically. This, by its very nature, implies statistical studies of large corpora. Nevertheless, the danger with such approaches, particularly the type of molecular-genetics approach adopted by Savage and his collaborators, is that they risk being focused on too low a descriptive level and may arrive at statistical generalities rather than meaningful particularities—the former an approach not dissimilar to the "beanbag genetics" criticized by Mayr (Dronamraju, 2010; but see also Juhász and Sipos, 2010). Savage and Atkinson (2015) concede this, arguing for the importance of taking into account:

> higher-level units of musical structure and meaning. In music, as in genetics, the individual notes that make up the sequences have little meaning in themselves. The phylogenetic analysis of sequences is thus merely the starting point from which to understand how and why these sequences combine to form higher-level functional units (e.g., motives, phrases) that co-evolve with their song texts and cultural contexts of music-making as they are passed down from singer to singer through centuries of oral tradition (Savage and Atkinson, 2015, p. 167).

**TABLE 1 |** Nature-culture mappings.

| Molecular-genetic | | Memetic | |
|---|---|---|---|
| **(Bio)chemistry** | **Music** | **Music** | **(Bio)chemistry** |
| Amino acid | Single pitch | Museme-element | Atom |
| ? | Motive | Museme | Molecule |
| Protein | Musical phrase | Museme sequence/ Musemeplex (see section 3) | Multi-molecule complex |

In this sense, it is important to consider—in the terms of the long-running debate in biology—the relevant *units of selection* (Lewontin, 1970), which requires a degree of nature-culture mapping[10]. While the protein sequences which Savage (2017) takes as analogous to musical sequences are useful exemplars of mutational operations, they have little evolutionary meaning in themselves. This is because *genes* are selected for, not nucleotides, nor, in Savage's case, the amino acids which make up the proteins whose production genes code for. Concomitantly, by focusing on discrete pitches—equated by Savage with the component amino acids of proteins—one is neglecting psychologically meaningful *groups* of pitches—these, in Savage's terms, equating to genes, which Dawkins regards as "any portion of chromosomal material that potentially lasts for enough generations to serve as a unit of natural selection" (Dawkins, 1989, p. 28). The mappings posited by Savage (2017) are summarized in **Table 1**, the first and second columns representing Savage's molecular-genetic mapping of (bio)chemical and musical structure, and the third and fourth columns representing a mirror-image, memetically motivated set of mappings (see also Jan, 2013, p. 152, Figure 1).

Thus, Savage's (2017) positing that amino acids are equivalent (in some abstract sense) to individual pitches and that proteins are equivalent to melodies is problematic because melodies are often made up of a number of discrete intermediate-level patterns—*musemes* (music-memes), in my terminology, and motives in Savage's (2017)—a crucial cognitive level which is not explicitly accounted for (hence the "?" in **Table 1**) in his approach. By "museme"—a particularly salient example of which is the opening four notes of Beethoven's Fifth Symphony—is meant a perceptually-cognitively-demarcated melodic/horizontal (pitch-rhythm) and/or harmonic/vertical collection which is capable of being retained in short-term memory and which possesses "just sufficient copying-fidelity to serve as a viable unit of [cultural] selection" (Dawkins, 1989, p. 195)[11].

[9]One might draw a distinction between formalization and quantitative studies: the former is an abstract attempt to theorize the terrain and dynamics of a system; the latter is a concrete attempt to measure a system using various metrics, perhaps using some formalization as a guide.

[10]These are not absolute correlations, but simply attempts to align phenomena at analogous structural levels within their parent "ontological category" (Velardo, 2016, p. 104, Figure 3).

[11]I alighted upon the term "museme" independently of Tagg (2016), conflating "music" and "meme" in an example of convergent evolution (*homoplasy*; see also

Such groups of pitches—the gene-equivalent patterns theorized by memetics—are much stronger candidates for the units of selection in cultural evolution than Savage's isolated pitches. This is because a m(us)eme is not a m(us)eme unless, as Dawkins states, it *can* act as a unit of selection. To serve this function it has to have a *discrete identity*; that is, it must (i) be discrete (demarcated to some extent from the patterns surrounding it, even if it partially overlaps with them; Jan, 2007, p. 74); and it must (ii) have an identity (it must have some attribute(s) which distinguish it to some extent from other, similarly demarcated, patterns and which motivate(s) its copying). These two points allow us to understand memetic selection as success in the competition for the finite attention and memory resources of a m(us)eme's potential human hosts.

There is very strong evidence from the cognitive-psychological literature that music is perceived in terms of such melodic/harmonic groups; and it would appear that they derive, in part, from the phenomenon of expectation (anticipation, prediction) (Huron, 2006; Husserl, 2013). As with many music-related perceptual-cognitive processes, this is a consequence of both bottom-up (innate/genetically determined) and top-down (learned/memetically determined) factors (Narmour, 1990). While subject to innate constraints, often considered under the rubric of Gestalt psychology, much of our perception of music (and indeed language) relies upon the statistical learning of conventions as a result of enculturation (Gjerdingen, 1988; Byros, 2009). This process has been modeled in a number of computer simulations: discussing their Information Dynamics of Music (*IDyOM*) model, Pearce and Wiggins (2012) argue that violation of expectations leads not only to affective responses (Meyer, 1956), but is a significant force in imposing grouping boundaries. Moreover, both bottom-up and top-down factors regulate the selective environment of musemes, for the former dictate the constraints a museme must satisfy in order to be perceived, cognized and memorized (Lerdahl, 1992; Velardo and Vallati, 2016); while the latter include the totality of musemes within a cultural community (the *museme-pool*), against which a given museme must compete (in the sense outlined at the end of the previous paragraph).

To expand upon the foregoing, one can make the following points:

- *Bottom-up:* Evolutionarily selected predispositions to vocal learning (Merker, 2012) make humans very good at attending to musilinguistic sounds (Brown, 2000; Mithen, 2006; Fitch, 2010) and abstracting statistical regularities from them (Kirby, 2013). This abstraction is fostered by the imposition of grouping boundaries, which "are perceived before events for which the unexpectedness of the outcome (*h*) and the uncertainty of the prediction (*H*) are high" (Pearce and Wiggins, 2012, p. 638). Such grouping boundaries create the "chunking" (Snyder, 2000, pp. 53–56) necessary for processing by short-term memory.
- *Top-down:* Suitably packaged, this musical information is retained in individual and collective memories; indeed, it would not be retained if it were not delineated. It might be termed, after Chomsky, "I-music" (internal, brain-stored, music) and "E-music" (external, culture-stored, music), respectively (Fitch, 2010, p. 32). Chunked musical patterns also influence the perception of other patterns, including their grouping, because "that which is copied [retained in memory] may serve to define the pattern" (Calvin, 1996, p. 21; see also Jan, 2011, section 4.1).

More broadly, the bottom-up/top-down duality raises the issue of gene-meme coevolution (Durham, 1991), because it pits biological replicators against their cultural equivalents. At the highest level, *system-orientated* research in coevolution encompasses the evolution of the human capacity for musicality and other phenotypic attributes (Blackmore, 2000, pp. 31–34; Jablonka and Lamb, 2014; Podlipniak, 2017); but *replicator-orientated* research in this area is generally not conducted with a specifically memetic orientation (but see Shennan, 2002), tending to focus on gene-level changes driven by (often generic) cultural pressures (Richerson et al., 2010). Thus, future research in coevolution might attempt to investigate meme-level changes driven by (specific) genetic pressures, and the interactions between specific memes and genes.

Given the foregoing, while the statistical data on folk-song corpora edits of Savage (2017) are strong evidence in favor of cultural evolution, they should be regarded as *epiphenomena* of musemic-evolutionary processes—consequences of the changes which occur when discrete musical patterns are transmitted with copying errors and are differentially selected. To gain a deeper understanding of such statistical data, one must regard the mutational changes (conservation, substitution, insertion and deletion) as forces not only *driving* musemic mutation and, ultimately, musico-stylistic evolution (Jan, 2015), but also as forces *constrained by* the psychological realities of pattern-formation and propagation. That is, one must take into account two countervailing forces: (i) susceptibility to mutational pressures (perhaps engendered by weak perceptual-cognitive demarcation and/or low intra-museme coherence) may distort a museme (resulting in high entropy; Margulis and Beatty, 2008), but may introduce a variant which has a higher perceptual-cognitive salience (Berlyne, 1971; Martindale, 1986), and therefore potentially greater replicative prospects, than its antecedent—Dawkins' "fecundity" (Dawkins, 1989, p. 194); and (ii) resistance to mutational pressures (perhaps engendered by strong perceptual-cognitive demarcation and/or high intra-museme coherence) may preserve multiple copies of a museme (resulting in low entropy), and may therefore foster an increase in its representation in the museme-pool over time—Dawkins' "copying-fidelity" (Dawkins, 1989, p. 195).

Lastly, one might argue that a memetic orientation erodes the qualitative-quantitative distinction—or, rather, that it allows us to understand it as a continuum—because it supports a range of methodologies from (qualitative) assessments of

section 3). While there are alignments between our uses of the term, mine is distinguished from Tagg's by its specifically evolutionary, as opposed to semiotic, focus—as a unit of cultural selection in music.

the aesthetic effects of certain musemes in particular musical contexts to (quantitative) measurements of museme frequency and transmission relationships.

## 3. QUANTIFICATION OF EVOLUTIONARY DISTANCE IN MUSEMES

To the first claim outlined in section 1—that a purely statistical approach based on counting note-edits without consideration of perceptual-cognitive aspects gives an incomplete account of cultural evolution—a second has arisen from section 2: that statistical data derived from measuring mutational changes, while illuminating, are epiphenomena of musemic evolution.

To investigate this, I consider some of Savage's (2017) specific data in a small case study, attempting to relate them to the musical patterns from which they arise. It is important to note at this stage that the tracking of conservations, substitutions, insertions and deletions is done partly in the service (in one of his studies) of grouping folk songs into tune-families (Cowdery, 1984), and I will focus on examples from one sub-family which will hopefully serve as a microcosm of more general issues. This focus is perhaps characteristic of the qualitative ("less is more")/quantitative ("more is more") distinction.

**Figure 1** shows one such melody, "The Two Brothers", no. 49 of the "Child Ballads", two variants of which are incorporated by Savage in his dataset. The Child Ballads are a collection of British folk ballads (specifically, their lyrics), assembled (some from American sources) by Child (1904). The (often diverse) melodies associated with these lyrics were later collated and categorized by Bronson (1959). This particular ballad, originally from Scotland, concerns the death—variously accidental or intentional—of one of the eponymous school-age brothers by the other's knife, and the deceased boy's subsequent interment[12].

What I label the "Antecedent" in **Figure 1A** (**Figure 1Bii**) was transcribed in Bronson's (1959) sources from a rendition by "Mrs. Ellie Johnson (23), Hot Springs, N.C., September 16, 1916" (Bronson, 1959, p. 391, no. 16); and the "Consequent" in **Figure 1A** (**Figure 1Biv**) from a rendition by "Mrs. Lucindie (G.K.) Freeman, Marion, N.C., September 3, 1918" (Bronson, 1959, p. 390, no. 15). Phrase-ending marks (represented by continuous vertical lines in **Figure 1B**) are Bronson's and are retained by Savage (2017). Being clear points of articulation, these marks are equivalent to the terminal nodes of the four musemes—Musemes (hereafter "M") *a*–*d*—which constitute these melodies (labeled under **Figure 1Biv**)[13]. While Savage is correct in labeling these two versions as "older" and "younger" (in terms of date of collection), respectively, there are actually four melodies in this group (six if one includes the variants in the second halves of two of them), and his "older" is not the "oldest": this status goes, by one day, to **Figure 1Bi**[14]. **Figure 1Bv**, represents the implied harmony of these melodies, which may or may not have been realized in some performances, perhaps on guitar.

While it makes sense methodologically for Savage (2017) to think in terms of "older" (antecedent) and "younger" (consequent) patterns, the fact that: (i) the time intervals between the recording of these phemotypic forms are so short (three days, in the case of **Figures 1Bi–iii**); (ii) the individuals concerned would presumably have assimilated these melodies months or years before the date of collection; and (iii) the geographical area from which they were collected is relatively constrained (the western counties of North Carolina, with two of the four melodies being collected in the same town, Hot Springs), all suggest that a model of linear transmission in collection-date order, with clearly demarcated, sequential mutations, is obviously highly improbable. This conclusion is further reinforced by the fact that the variants in **Figures 1Bi,iii**, were presumably recorded on the same occasions as the ostensibly "principal" form. Given these points, references in what follows to "earlier"/"antecedent" and "later"/"consequent" forms of melodies and musemes must be understood as relating only to the dates of collection and to the resultant numeration in **Figure 1B**, and not as hypotheses of evolutionary descent-order.

An arguably more realistic model would be of an ecosystem in which a relatively stable framework—defined by balanced and rhyming periodicity, implied harmony, cadence patterns and axial pitches—was generated by means of a number of interchangeable musemes being repeatedly co-replicated. This framework is eight bars in duration, with a I–V; V–I two-phrase/four-sub-phrase structure and a "middle cadence …on the supertonic [$\hat{2}$, supported by an implied V]" (Bronson, 1959, p. 384). It is clearly not unique to this set of song variants: it forms the basis, much expanded, of "two-phrase"/"balanced" binary form (Rosen, 1988, p. 22; Hepokoski and Darcy, 2006, p. 355), as well as of numerous other folk-song melodies (Bronson, 1959, p. xii)[15]. It serves as a container for a set of musemes which were interchangeable in ways which did not compromise the integrity of the melody, as understood by members of the cultural community which replicated it in conjunction with a similarly variable set of verbal-conceptual (lyric/text) memes.

In this sense, "The Two Brothers" is a higher-order structure re-instantiated/generated by the repeated re-conglomeration of a set of functionally equivalent musemes, each of which serves to articulate a specific node of the structure. The notion of functionally analogous musemes is essentially that of the replicator *allele* (Dawkins, 1983a, p. 283). This concept, when used in the context of cultural evolution, refers to musemes which are similar in their basic structure and/or function, such that members of the same museme *allele-class*

[12]Verses 4 and 6 of one variant of this ballad read: "4: Brother took out his little penknife, / It was sharp and keen. / He stuck it in his own brother's heart, / It caused a deadly wound. 6: He buried his bible at his head, / His hymn book at his feet, / His bow and arrow by his side, / And now he's fast asleep." (Bronson, 1959, p. 391).

[13]The segmentation of these melodies is largely unproblematic, being guided by Gestalt-psychological segmentation criteria (Deutsch, 1999; Snyder, 2000). While a phrase is not necessarily the same as a museme, in the case of this melody the four short phrases are. The distinctive ♩.–♪ rhythm straddling M*c*–M*d* in most of these melodies also acts (residually) in the two examples where the junction is ♩–♩, i.e., the variant forms of **Figures 1Bi,iii.**

[14]While Bronson categorizes these six as belonging to "Group B" of this tune-collection (Bronson, 1959, pp. 387–393, nos. 9–20), others in this group are often significantly different to the homogeneous six which are shown in **Figure 1B.**

[15]Such similarities suggests a deep commonality between song and dance melodies arising from the imperatives of symmetry, balance and an arch-shaped (low-high-low) tension-curve.

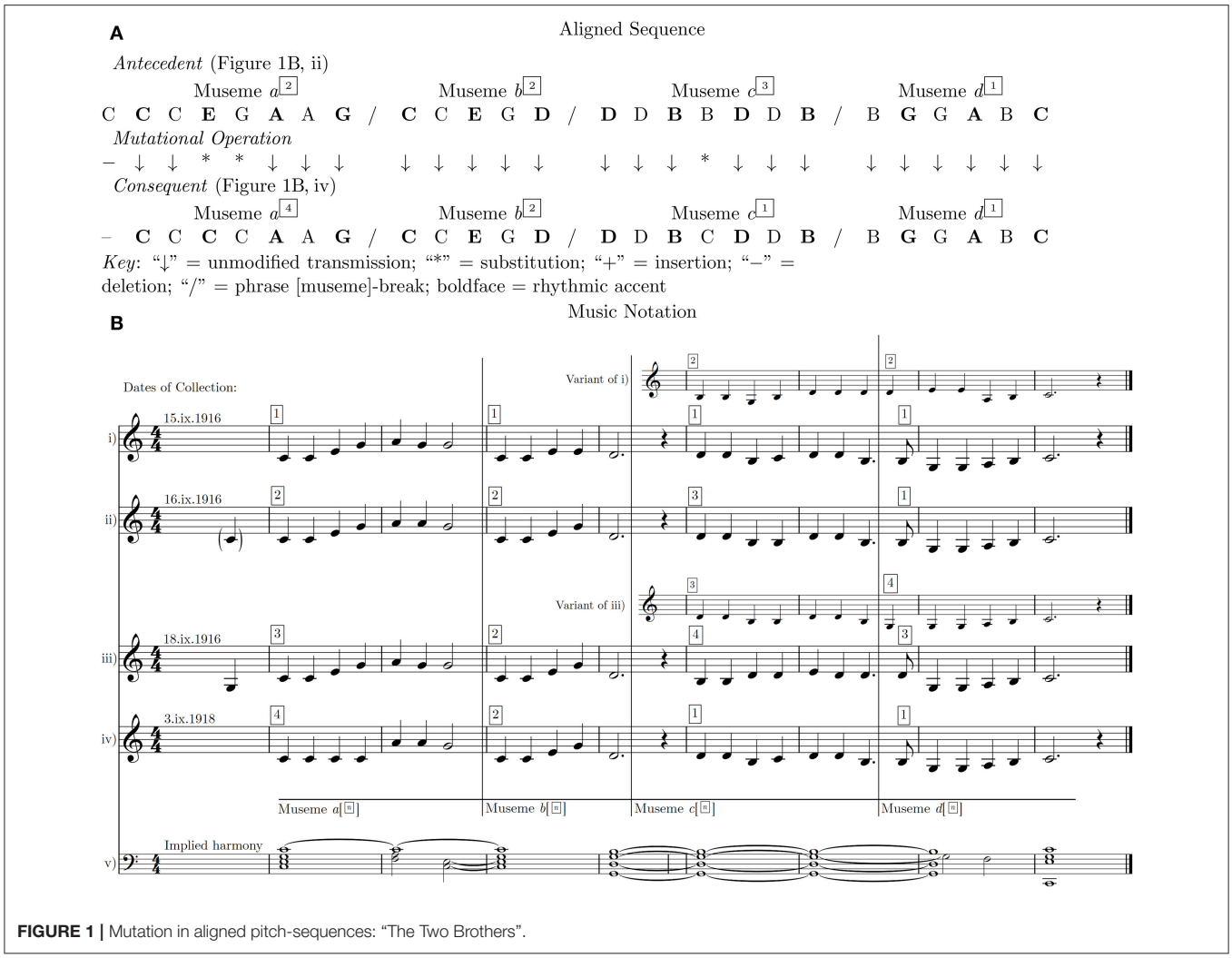



**FIGURE 1 |** Mutation in aligned pitch-sequences: "The Two Brothers".

are interchangeable—in the sense of being equally viable and coherent—in a specific context (such as a certain point in a phrase or a particular modulatory juncture, etc.) (Jan, 2016). The framework/higher-order structure referred to above might be termed a *musemeplex*—i.e., a complex formed by the repeated co-replication of a set of musemes (in the case of **Figure 1B**, M*a*–M*d*) which are nevertheless also individually replicated (Jan, 2007, p. 80). Automatically, the replication of a musemeplex results in the replication of what might be termed a *musemesatz*—i.e., a shallow-middleground-level structure, the "skeleton" of a musemeplex, generated by the tendency of a set of allelically related musemes to conglomerate in broadly similar ways in two or more contexts (Jan, 2010). As represented in **Figure 1**, allele-identifiers are shown as superscript boxed Arabic numbers (assigned according to date of collection), so that (for example) bb. 1–2 of **Figure 1Bi**, is labeled Allele 1 of M*a*, symbolized hereafter in the text as "M*a*[1]"[16].

Given this nexus of similarity relationships linking six melodies assembled from a set of fourteen alleles, how might we understand the connections between the component musemes and attempt to reconstruct their transmission relationships? Perhaps it is necessary to concede that one cannot ultimately reconstruct the nexus of transmission that gave rise to these six melodic variants, simply because human culture is so interconnected—and was even when these songs were current, in the pre-internet age—and the cultural interactions with which we are concerned were largely undocumented. But one might still try to sketch out possible evolutionary trajectories and develop methodologies which might be applicable to these and other cases. One way is to attempt to quantify the differences between them, in terms of measuring the mutational changes that separate them. Savage proposes the percent identity (PID) as a measure of evolutionary distance, this being defined as "the

[16]Such numeration is, naturally, undertaken "vertically" (i.e., as an *intra*-museme-allele-class system), and not "horizontally" (i.e., as an *inter*-museme system). The latter approach would indicate a degree of similarity, for example, between M*a*[3] and M*c*[4] (**Figure 1Biii**). Nevertheless, this particular connection is a *contour*-based similarity, and not one inhering in the *intervallic* and *scale-degree* recurrences which I employ here to define similarities and differences between musemes and their alleles.

number of aligned positions (i.e., amino acids, DNA nucleotides, musical notes, etc.) that are identical (ID) divided by the sequence length (L)…. We have chosen to use the average length of both sequences [$L_1$, $L_2$], as this appears to be the most consistent measure of percent identity" (Savage, 2017, pp. 53–54). This metric is represented in the following equation:[17]

$$PID = 100 \times \{\frac{ID}{\frac{L_1+L_2}{2}}\} \quad (1)$$

Savage (2017) uses the PID as an index of the *mutational distance* between two variant *melodies* in order to assess a tune's membership of a particular tune-family—the larger the PID, the greater the likelihood of the melodies' belonging in the same tune-family. But there is no reason why this metric cannot also be used at the level of the *museme*, in order to quantify mutational distance between such patterns. Used this way, the PID may be used to assess membership of a museme allele-class (or, indeed, to investigate a relationship of presumed mutation which moves a museme from one allele-class into another). Membership of a museme allele-class implies—provided the musemes are of a comparable length—that the musemes in question are related by homology ["a character shared between two or more species that was present in their common ancestor" (Ridley, 2004, pp. 427, 480); what Darwin termed "descent with modification" (Darwin, 2008, p. 129)], rather than homoplasy ["a character shared between two or more species that was not present in their common ancestor" (Ridley, 2004, pp. 427–428, 480)]; that is, a relationship resulting from cultural transmission, rather than from "convergent evolution" (Ridley, 2004, p. 429), respectively. Nevertheless, as with comparable cases in biology, it is not always possible to decide with certainty which category specific cases belong in. While determination of a suitable PID threshold for perceptually-cognitively significant similarity might be achieved by means of empirical studies—whereby test musemes with various degrees of mutation are ranked by listeners according to their perceived relatedness—this would not necessarily permit the assignment of threshold-exceeding patterns to the same allele-class without fuller knowledge of the context of transmission.

A related metric is *mutation rate*, which is the number of "observed mutations per year" (Savage, 2017, p. 56), where the number of mutated pitches ($x$) is compared with the total number of pitches ($y$) over time ($t$). This is represented in the following equation:

$$MR_{(t)} = (x/y)/t \quad (2)$$

Again, there is no reason why this metric cannot also be used at the museme level, in order to quantify the mutation rate between two museme alleles. While cultural evolution occurs at an absolute rate many orders of magnitude faster than biological evolution (Dawkins, 1989, p. 192), and indeed occurs at highly variable absolute rates (Savage, 2017, p. 107), if cultural evolution is scaled to biological evolution (i.e., if some relative rather than absolute mutation rate is considered), then the two processes may be broadly comparable. Mutation rate is directly correlated with "transmission fidelity" (Savage, 2017, p. 111), in that the lowest mutation rates are found in repertoires with high copying-fidelity, and *vice versa* (Dawkins, 1989, pp. 18, 194); these repertories tend, unsurprisingly, to be notationally (as opposed to orally) transmitted musics. In the case of these particular melodies, however, the time interval is so constrained, and the transmission nexus sufficiently unclear, for the mutation-rate metric to be of limited use (despite the illustrative calculation below) in the present context.

On this basis, the PID and MR values (the latter over a notional 2-year period, the time interval separating the collection of **Figures 1Bii,iv**) for M$a^{[2]}$ and M$a^{[4]}$ in **Figure 1** are as follows:

$$PID = 100 \times \{\frac{5}{\frac{8+7}{2}}\} = 71.4 \quad (3)$$

$$MR_{(t)} = (3/8)/2 = 0.188 \quad (4)$$

Because the musemes under investigation are components of a larger melody—they are, as argued above, independently replicated elements of a musemeplex which is transmitted, iso-sequentially ordered, as a collective—when the melody is copied from source to source, it is clear that the order and identity of musemes is either retained or obviously altered[18]. Such cases of musemic transmission are therefore more tractable—M$a^{[2]}$ in one melody is clearly analogous to M$a^{[4]}$ in a variant of that melody—than situations in which an isolated museme is potentially copied from an antecedent context (a piano sonata, for example) to a non-analogous consequent context (a symphony, for example). In the latter case, however, the PID and MR metrics might usefully be employed in order to assess the likelihood that a given pattern is indeed being transmitted from one context to another.

Such sequential-mapping constraints allow one to circumvent the fact that, at 71.4%, the PID value of M$a^{[2]}$–M$a^{[4]}$ in **Figures 1Bii,iv** is lower than the 85% Savage takes as an index of two melodies being "highly related" (Savage, 2017, p. 54)[19]. It is conceivable, however, that two melodies with a PID of this order of magnitude may not actually bear any obvious musemic relationships, owing to the insensitivity of the PID metric to museme similarity when the PID is calculated at the musemeplex (phrase) level (one might address this by calculating the PID at the musemeplex level using musemes rather than individual pitches as the units of measurement)[20]. Because Savage's (2017)

[17] While Savage (2017, p. 51) argues for, and operationalises, the primacy of pitch over rhythm in his melodic-similarity determinations—yet usefully takes into account the distinction between accented and unaccented pitches—future research in this area might usefully integrate both parameters in a more sophisticated PID metric.

[18] This attribute of independent replication is assumed for the sake of argument, but it is not difficult to envisage easily finding *coindexes* (Jan, 2007, p. 71) of the individual musemes of "The Two Brothers", replicated separately from the assemblage of which they form a part in the ballad.

[19] A PID <85% may still indicate a relationship of (partial) transmission, in which one or more musemes from one melody are assimilated by another, largely dissimilar, melody.

[20] This is a consequence of the phenomenon famously summed up by the comedian Eric Morecambe, who said to André Previn—after a shambolic start by Morecambe to Grieg's Piano Concerto in A minor—"I'm playing all the right notes—but not necessarily in the right order" (McCann, 1999, p. 234).

≥85% criterion applies to melodies, not musemes, and because his algorithm has paired the 71.4%-related M*a*[2] and M*a*[4] in **Figures 1Aii,iv**, there must by definition be a >85% similarity between the other musemes of the phrase, M*b*[n]–M*d*[n], in order to compensate for the <85% of the M*a*[2]–M*a*[4] relationship. Indeed, M*b*[2] and M*d*[1] are replicated (as their symbology implies) without mutation (= 100% relation).

**Table 2** shows PID values for each museme allele-class in "The Two Brothers", comparing alleles of M*a*–M*d* against others in the same allele-class[21].

Without the anchor of the sequential-mapping constraint, many of these patterns would not, on the basis of their PID values, appear to be related. The similarities between M*a*[2] and M*a*[4], for example, inhere in relatively tenuous pitch connections—the 28.6% "PnID" (Percent *non*-IDentity = 100%−71.4) puts quite an expanse of clear blue water between them. In the case of the M*c*[1]–M*c*[4] relationship, the considerably smaller 14.3% PID value (and therefore considerably greater 85.7% PnID) would not even suggest membership of the same allele-class[22]. In both cases, and as is often the case in musemic similarity relationships, it is the rhythm, contour and harmonic implication—the latter a prolongation of the tonic and dominant chords, respectively (**Figure 1Bv**)—which additionally binds these alleles together (and which would have to suffice in the absence of the sequential-mapping constraint). In the case of M*a*[2] and M*a*[4], the rise from the initial $c^1$ to the apical $a^1$ in b. 2 followed by a fall to the dominant $g^1$ at the end of the first half-phrase is the common, unifying contour feature of the allele-class.

Measures of similarity have a bearing on the related issues of museme transmission and of museme resolution/subdivision. In general, cultural transmission is significantly more error-prone (in an informational sense) than biological transmission, so it may be presumed that most inter-museme PID values will be lower than 100%[23]. Below a certain context-specific threshold, a low PID value might be taken as evidence that any similarities are the consequences of homoplasy, not homology. But the converse may not always hold true: a very high PID might be associated with a pattern so generic and so commonplace that the two instances may have been independently generated (homoplasy), rather than directly transmitted (homology). In Cope's terms, such entities are "commonalities": a category of "patterns which, by virtue of their simplicity—scales, triad outlines, and so on—appear everywhere. In a sense, commonalities seem to disappear in a sea of similarity" (Cope, 2003, p. 17). By contrast, and at the opposite end of a continuum of similarity categories (Jan, 2014, p. 4, Figure 1), longer and more distinctive patterns are termed "quotations": a category which "often involve exact note and/or rhythm duplication" (Cope, 2003, p. 11). Quotations are more likely than commonalities to be homologous as opposed to homoplasious, and *vice versa*. Thus, one must also take into consideration the issue of museme length, in addition to the PID value, when attempting to determine whether two coindexes are related by homology or by homoplasy.

On this last point, and as noted in section 2, museme perception and cognition is contingent upon both bottom-up and top-down processing. The former to some extent tracks the sonic-acoustic regularities governed by the laws of physics. Given that these regularities include the harmonic series, it is perhaps not surprising that certain musical structures derived from this series—triads and particular (5–7-note, unequal-interval) scale-types—are common across many (but not all) musical cultures (Patel, 2008, pp. 19–21). Such structures are thus to some extent acoustically privileged and will (*ceteris paribus*) naturally constitute the "connective tissue", the commonalities, of much music—which is not to say that the particular (rhythmic/harmonic) form they take in a given piece of music is not derived (memetically) from a specific antecedent coindex. Moreover, such commonalities are often useful in expediting the connection of more "characteristic" musemes (i.e., those closer to the "quotations" end than the "commonalities" end of Cope's (2003) continuum) and, in this capacity, they therefore serve as evolutionary "good tricks" (Dennett, 1995, pp. 77–78).

As a further complication, similarity values are often not helpful in trying to order musemes chronologically/sequentially in a nexus of transmission. As will be discussed further in section 4, evolution is not invariably associated with increasing complexity, however measured; in certain circumstances, adaptation might result in decreasing complexity. Moreover, the PID value measures editorial *differences* (it is not, strictly, an edit-distance metric; Levenshtein, 1966), which might result in no net change in absolute or relative complexity between two or more musemes; nor does it indicate the *direction* of change (toward greater simplicity or greater complexity), so a high PID might be associated with operations which result in the simplification of a museme, such as occurs between M*a*[2] and M*a*[4]. Of course, this relationship is only one of simplification if M*a*[2] is regarded as the antecedent and M*a*[4] as the consequent; seen the other way round, it is a process of increasing complexity. If evolution were *only* taken to be a process of increasing complexity, then M*a*[4] would be a candidate for the antecedent of M*a*[2]—which it might nevertheless still be, even though this specific (simplicity-complexity) justification is invalid.

Hitherto, these alleles have been treated as unitary, but if we hypothesize that three notes is the realistic lower threshold for a melodic museme to have perceptual-cognitive validity (Jan, 2007, p. 61), then the $a^1$–$a^1$–$g^1$ melodic triad of b. 2 is the only common contiguous element between M*a*[2] and M*a*[4]. (One might, however, regard Musemes M*a*[2] and M*a*[4] as being identical at the shallow-middleground level—having a $c^1$–$a^1$–$g^1$

[21]The bracketed anacrusis $c^1$ (Bronson, 1959, p. 391, no. 16) in **Figure 1Bii** is included here, as it is in Savage's mutation calculation, represented in **Figure 1A**.

[22]Perhaps criteria might be devised which would conclude that they are not actually in the same allele-class, or that they are only members of an "allele-super-class", perhaps one defined by harmony but not including scale-degree factors. While the present focus is largely upon melodic (linear pitch plus rhythm) patterning, one could vary the number of parameters taken into consideration in order to narrow or broaden the definition of a museme. In this way, a museme would be seen as a multiparametric complex (a "style structure", in Narmour's terminology) made up of several uniparametric simplexes ("style forms/shapes") (Narmour, 1977, pp. 173–174; 1990, p. 34), although this runs the risk of blurring the distinction, if one truly exists, between a museme and a musemeplex.

[23]One concomitant of the dichotomy expressed in Note 5 is that a work-centric view of music attempted, until relatively recently, to enforce a single correct and objective text, whereas a process-centric view accepts the diversity of different acts, be these interpretations of "classical" works or variants of folk musics.

**TABLE 2 |** PID values for museme alleles in "The Two Brothers".

| | $Ma^{1}$ | $Ma^{2}$ | $Ma^{3}$ | $Ma^{4}$ | $Mb^{1}$ | $Mb^{2}$ | $Mc^{1}$ | $Mc^{2}$ | $Mc^{3}$ | $Mc^{4}$ | $Md^{1}$ | $Md^{2}$ | $Md^{3}$ | $Md^{4}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Ma^{1}$ | | 85.7 | 100 | 57.1 | | 66.7 | | | | | | | | |
| $Ma^{2}$ | | | 75.0 | 71.4 | | | | | | | | | | |
| $Ma^{3}$ | | | | 57.1 | | | | | | | | | | |
| $Ma^{4}$ | | | | | | | | | | | | | | |
| $Mb^{1}$ | | | | | | 80 | | | | | | | | |
| $Mb^{2}$ | | | | | | | | | | | | | | |
| $Mc^{1}$ | | | | | | | | 28.6 | 85.7 | 14.3 | | | | |
| $Mc^{2}$ | | | | | | | | | 42.9 | 57.1 | | | | |
| $Mc^{3}$ | | | | | | | | | | 14.3 | | | | |
| $Mc^{4}$ | | | | | | | | | | | | | | |
| $Md^{1}$ | | | | | | | | | | | | 50 | 83.3 | 83.3 |
| $Md^{2}$ | | | | | | | | | | | | | 83.3 | 50 |
| $Md^{3}$ | | | | | | | | | | | | | | 83.3 |
| $Md^{4}$ | | | | | | | | | | | | | | |

structure; but a full consideration of the structural-hierarchic location of the musemes under consideration is beyond the scope of the present article.) The first part of the museme—$(c^1)$–$c^1$–$c^1$–$e^1$–$g^1$ in $Ma^{2}$, []–$c^1$–$c^1$–$c^1$–$c^1$ in $Ma^{4}$—is sufficiently dissimilar (despite the two common $c^1$s) for one to envisage various scenarios to account for the etiology of the material of bb. 1–2 in these two song-variants, scenarios which may be generalized to other musemes in these six melodies and, indeed, more widely.

To contextualize these scenarios, it is useful to make a distinction between two ways of viewing these melodies and the alleles which constitute them, which might be conceived as extreme points on a "continuum of influence". On the one hand (the imaginary left-hand ("closed") side of the continuum), one could see these six melodies as an essentially secure ecosystem, impervious to perturbation by musemes external to its constituent allele classes. On the other hand (the imaginary right-hand ("open") side of the continuum), one could see them as entirely receptive to influence by external factors (immigration of, or influence by, external musemes). In the case of "The Two Brothers", it seems sensible to ascribe priority to intra-tune-family relationships, given the nature of this repertoire's transmission, while not ruling out the possibility that musemes from other sources—other tune-families, other repertoires—might have influenced the transmission relationships within this group of six melodic variants. It is also important to note that in such repertoires as the folk ballad there is obviously textual as well as musical replication, but this does not necessarily guarantee that, when a textual phrase is replicated from one context to another, the museme associated with the earlier text is the source of that associated with the later text—as other instances of "The Two Brothers" tune-family attest.

For $Ma^{n}$ and the multitude of comparable cases:

1. One could regard bb. 1–2 of "The Two Brothers" as consisting of only *one* museme ($Ma^{2}$ and $Ma^{4}$). If so, then given the similarities between the second halves of each variant (the $a^1$–$a^1$–$g^1$ triad), which act as a kind of "anchor" (and given, of course, the sequential-mapping constraint), one would take the first halves, b. 1, as being edit-heavy, homology-associated mutations: to get from the antecedent to the consequent form (whichever is which), a fair amount of "earth moving" is required (Typke et al., 2007; see also Jan, 2014).
2. Alternatively, as shown in **Figure 2**, one could regard bb. 1–2 as consisting of *two* musemes (or two museme allele-classes), which one might label $Ma^{2}_{x}$/$Ma^{4}_{x}$ and $Ma^{2+4}_{y}$ (the latter being the $a^1$–$a^1$–$g^1$ triad). Under this alternative interpretation, then:
   (a) Liberated from their evolutionary dependency with $Ma^{2+4}_{y}$, the relationship between $Ma^{2}_{x}$ and $Ma^{4}_{x}$ could be one of either homology (i.e., $Ma^{4}_{x}$ is copied from $Ma^{2}_{x}$ (or *vice versa*)) or homoplasy (i.e., $Ma^{4}_{x}$ is copied from an antecedent other than $Ma^{2}_{x}$). Homology might be more likely to be the case if one were situated on the left-hand/closed side of the "continuum of influence" referred to above; and homoplasy might be more likely to be the case if one were situated on the right-hand/open side of the continuum.
   (b) Given its relative brevity, the same qualification as to homology *versus* homoplasy applies to $Ma^{2+4}_{y}$, which is a commonality (in Cope's (2003) terms) of tonal music. Thus, while perhaps unlikely on account of the wider melodic similarities, it could in principle be the case that both $Ma^{2\text{ or }4}_{x}$ and $Ma^{2+4}_{y}$ are separately transmitted to the consequent of "The Two Brothers", circumventing the posited antecedent.
3. For all these scenarios, some degree of blending inheritance might have occurred: positioned in the center of the continuum of influence, an intra-tune-family transmission event might have been influenced by an extra-tune-family factor. Thus, if **Figure 1Bii** were antecedent to **Figure 1Biv**

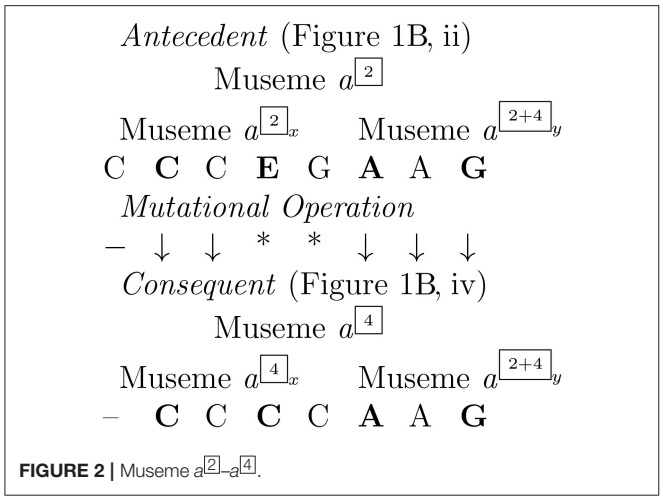



**FIGURE 2 |** Museme $a^{[2]}$–$a^{[4]}$.

then replication of the latter might have been meditated by the memory of a melody containing a repeated-note museme.

Given that **Table 2** shows intra-museme-allele-class PID values, what is not considered are inter-museme-allele-class values. One of the latter is, however, shown (italicized), namely that between M$a^{[1]}$ and M$b^{[2]}$, the relatively high value of 66.7% (higher, of course, than some intra-museme-allele-class values) indicating the presence of rhyme/symmetry within the first half of the melody[24]. The higher the intra-museme-allele-class ("vertical") PID values of any tune-family, the greater the perceived *synchronic* unity (its coherence as a collection of melodies) of the family; whereas the higher the inter-museme-allele-class ("horizontal") values of any individual melody, the greater the perceived *diachronic* unity (its coherence as a collection of musemes) of that melody—and *vice versa*. Both forms of unity might act as musemic selection pressures: the higher the perceived unity, synchronic or diachronic, the easier it is for listeners and singers to remember these melodies and therefore the more evolutionarily successful their constituent musemes may tend to be, if success is measured in terms of the number of copies of a given museme in a museme-pool. This selection pressure might be operative in many musemeplexes, and might be a factor driving the musemic collaboration which gives rise to them.

## 4. PHYLOMEMETICS AND CULTURAL TAXONOMIES

The reference to "phylogenetic analysis" in the quotation in section 2 (page 3) is significant, in that just as the long-term outcomes of biological selection can be represented in terms of branching lineages on (by convention) a tree diagram—where species bifurcate to give rise to sub-species, etc. (Darwin, 2008, p. 90)—so can those of cultural evolution. In the case of the group of museme alleles constituting the particular subset of "The Two Brothers" tune-family shown in **Figure 1**, one might apply the principles of cladistic taxonomy (Hennig, 1999) to arrive at a representation, a *cladogram*, not of the evolutionary relationships between "dialects" (arguably the cultural equivalent of species; Meyer, 1996, p. 23), but between musemes (the cultural equivalent of genes)[25]. Thus, this enterprise is closer to molecular genetics than it is to species taxonomy.

As a first word of caution, attempting to calculate cultural phylogenies—what might be termed *phylomemies*—from such a small group of short melodies risks falling foul of what might be termed the distinction between *real* and *virtual* phylogen/memies. A real phylogen/memy is one which is objectively evolutionarily correct, indicating the transmission relationships between the replicators at various positions on the cladogram. A virtual phylogen/memy is one which arrives (perhaps as a consequence of a restricted sample size) at a "pseudo-cladogram" which, while a logical and (perhaps more importantly) parsimonious representation of the patterns under investigation, is nevertheless (potentially) *not* evolutionarily true (and is therefore not properly cladistic) because it does not take into account patterning "external" to the sample under consideration. This external patterning, if included, might alter the relationships represented by the cladogram. It would appear considerably easier to arrive at a real phylogeny (where groups of potentially related organisms are often relatively geographically localized, morphologically distinct and, nowadays, genetically tractable) than it is to arrive at a real phylomemy (where groups of potentially related cultural forms are often scattered across space and time).

Yet this enterprise is worth pursuing, if only to illustrate the possibilities of the approach, one which Howe and Windram (2011) term "phylomemetics", the cultural equivalent of phylogenetics. As they acknowledge (Howe and Windram, 2011, p. 1), this is by no means a new methodology in the humanities, where philologists in both linguistic and musical research have long attempted to reconstruct *stemmata* showing relationships of transmission and mutation in sources as diverse as biblical texts and medieval music manuscripts (Cook, 2015). Conducted under (or, some might fear, annexed by) the rubric of phylomemetics, such research can incorporate all the intellectual infrastructure of Darwinism—the notions of variation, replication and selection; concepts of fitness; and ideas of lineage bifurcation and divergence—in tracing connections between the phenomena under investigation[26].

Using the phylogeny-calculation software *Phylip* (Felsenstein, 2016), the six forms of "The Two Brothers" in **Figure 1B** were

[24]Bronson argues for the primacy of musical over textual rhyme (Bronson, 1959, p. xii).

[25]There are various different approaches to taxonomy, and biologists often argue testily as to their relative merits—in Dawkins' view, taxonomy is "one of the most rancorously ill-tempered of biological fields. Stephen [Jay] Gould has well characterized it with the phrase 'names and nastiness' " (Dawkins, 2006, p. 275). But a cladistic approach, particularly one where genetic evidence is employed, is the one most likely to be evolutionarily "correct" in biological taxonomy (Ridley, 2004, p. 489).

[26]It might be argued that phylomemies differ from phylogenies in their potential for "cross-fertilization", whereby two lineages may share material, or even rejoin, after bifurcation. But this is also true, perhaps to a lesser extent, in nature, where gene-transfer between recently bifurcated lineages remains possible for a limited time.

**A** For Figure 1B, i–iv

```
6  26
15sept1916 0ccegagg cceed ddbcddb bggabc
15sept16-v 0ccegagg cceed bbgbddd deeabc
16sept1916 cccegaag ccegd ddbbddb bggabc
18sept1916 gccegagg ccegd bbddedd dggabc
18sept16-v gccegagg ccegd ddbbddb gggabc
03sept1918 0ccccaag ccegd ddbcddb bggabc
```

**B** For Figure 1B, i–iv, Museme *c* Only

```
6  7
15sept1916 ddbcddb
15sept16-v bbgbddd
16sept1916 ddbbddb
18sept1916 bbddedd
18sept16-v ddbbddb
03sept1918 ddbcddb
```

**FIGURE 3 |** Input data for "The Two Brothers".

analyzed. This used the input file shown in **Figure 3A**, which is a date-ordered list—based on **Figure 1B** and in which "v" represents the variant forms of **Figures 1Bi,iii**—of the melodies consisting of a sequence of their constituent pitches, grouped into museme alleles[27]. It should be stressed that this is an illustrative calculation only, designed to outline a methodology which might be adopted (as discussed in section 5) in larger studies. The highly restricted dataset naturally limits the scope of the conclusions that can be drawn. The phylomemetic tree shown in **Figure 4A** was generated using the *Pars* utility, which "is a general parsimony program which carries out the Wagner parsimony method (Eck and Dayhoff, 1966) with multiple states. Wagner parsimony allows changes among all states. The criterion is to find the tree which requires the minimum number of changes" (Felsenstein, 2016). For ease of comparison, the text-based output of *Pars* (strictly, that of the graphics-generating utility *DrawGram*) has been replaced in **Figure 4** by images of the relevant melodies[28].

Such cladograms represent descent with modification, whereby items located to the left (bottom/past) are hypothesized to be evolutionarily earlier than those located to the right (top/present), and where proximity to points of bifurcation (branch-length) represents relative evolutionary distance. While parsimony does not invariably align with evolutionary reality (a parsimonious tree is not necessarily a "real" tree, in terms of the binarism referred to above), it is a powerful constraint on evolutionary possibilities. Given this, it is reasonable to infer that both real and virtual lineages will generally proceed from left to right by the minimal mutational distances (this is not to deny the possibility of more radical, saltational, change). As suggested in section 3, evolution is fundamentally a process of *adaptive change* (Ridley, 2004, p. 4) and not necessarily one where that change leads to an increase in "the logarithm of the total information content of the biosystem (genes plus memes)" (Ball, 1984, p. 154)[29]. In the light of this, and of the proviso made in section 3 that date of collection does not necessarily align with the evolutionary chronology of these melodies, one must reiterate that, when undertaking phylomemetic analysis, melodic simplicity does not necessarily correlate with chronological anteriority, any more than melodic complexity corresponds with chronological posteriority.

As a second word of caution—one which applies more broadly to any attempt to analyse music by means of the kinds of symbolic representations used in *Phylip*—in order to perform the phylomemetic analysis, the musical patterning of these songs, already converted to their traditional western letter-name notation in **Figure 1**, was rendered as a series of ASCII characters to form the input to *Pars*. In this way, the melodies of these ballads are treated as a text. This means that the analysis is operating on a representation two stages removed from the living performances recorded over a century ago: not only has the rendition been regularized and shoehorned into western notation, a form of "lossy" compression; but this representation has itself been further divorced from its connection with sound by its reduction to a mere symbol-set, an abstract series of M$x^{\boxed{n}}$ patterns. Perhaps more fundamentally, while the *Phylip* software to some extent "understands" genetics, in that it is based on a formalization of the dynamics of the biochemistry underpinning it, it has little conception of music and the dynamics of pitch and rhythm combination underpinning it. Nevertheless, the symbols offered as input bear at least some connection with their long-distant musical antecedents, and so permit a provisional phylomemetic analysis based on parsimony relationships to be conducted.

In addition to analyzing relationships between song melodies as a whole, this type of analysis may also be conducted at the level of the museme allele, as represented in **Figures 3B**, **4B**, which show only the four alleles of M*c*. Importantly, if cladograms generated from complete song melodies are different from those derived from specific museme alleles within a melody, then this affords evidence in support of the second claim, made in section 3: that statistical data derived from measuring mutational changes, while illuminating, are epiphenomena of musemic evolution.

While there are many complex relationships represented within the cladograms of **Figure 4**, not all of which can be elaborated upon here, the following points may be made in summary (again reiterating that the *Pars* utility is operating on a deprecated, symbolic representation of music without any knowledge of music theory):

[27]This might be further developed by incorporating rhythmic values, whereby "bbb" = ♩. and "b" = ♪.

[28]Note that these are "rooted" phylomemies: there is assumed to be an unidentified common ancestor to the left of the tree (Ridley, 2004, p. 439).

[29]This may often be the case with oral transmission, where the principle of *lectio difficilior potior*—"the more difficult reading is the stronger" (Robinson, 2001)—might support one in ascribing chronological anteriority to a more complex form.

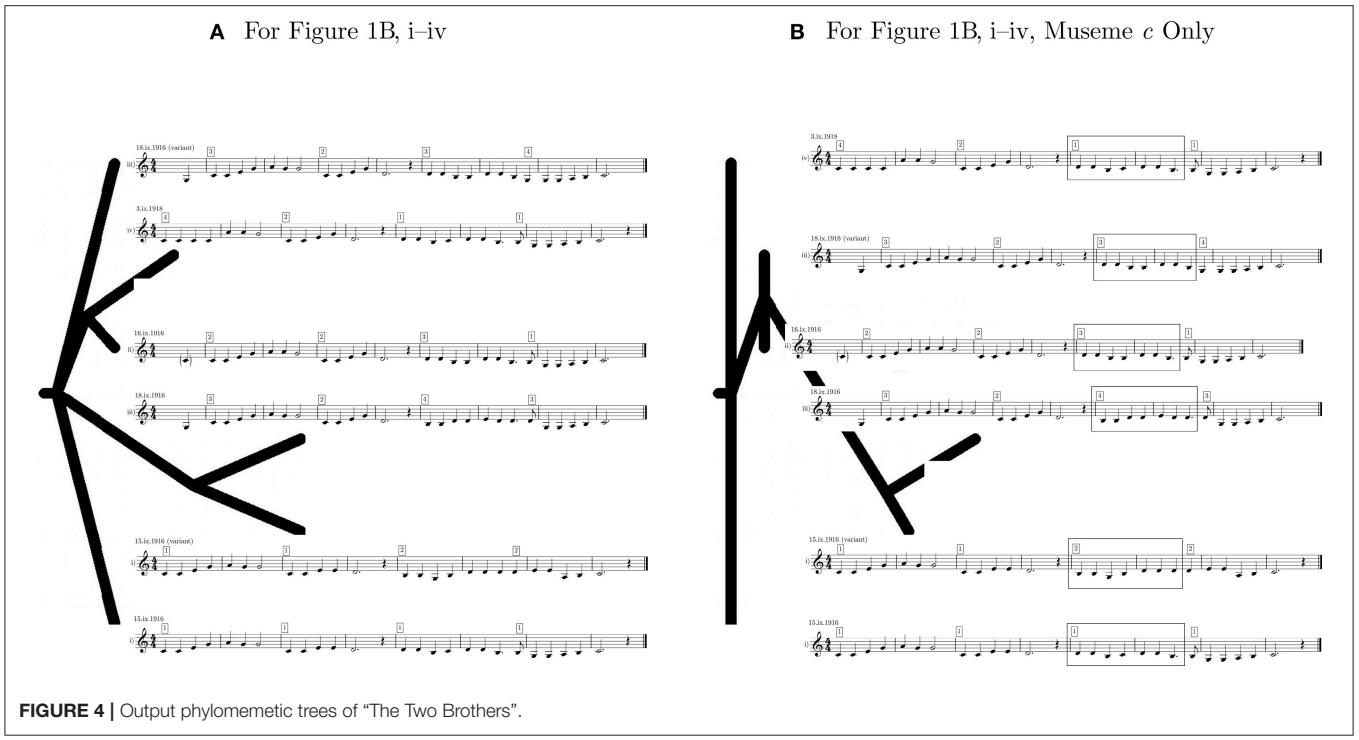



**FIGURE 4 |** Output phylomemetic trees of "The Two Brothers".

1. In **Figure 4A**, the melodies shown in **Figures 1Bi,iii** (variant) are hypothesized to be evolutionarily prior and are distinguished by the difference between M*b*[1] and M*b*[2] and by a pitch difference between M*c*[1] and M*c*[3].
2. In the same cladogram, two groupings of posited evolutionary descendants link **Figures 1Bii,iv** (perhaps by virtue of the common $a^1$–$\underline{a}^1$–$g^1$ melodic triad in M*a*[2] and M*a*[4] (designated earlier as M$a^{[2+4]_y}$)); and **Figure 1Bi** (variant) and **Figure 1Biii** (perhaps by virtue of the common $a^1$–$\underline{g}^1$–$g^1$ melodic triad in M*a*[1] and M*a*[3] (which might, by extension with M$a^{[2+4]_y}$, be designated M$a^{[1+3]_y}$), and (in the same pair) of the prominence of the pitch $d^1$ toward the end of M*c*[2] and M*c*[4]).
3. In terms of chronology, this first cladogram broadly aligns with the dates of collection of these songs; but, as noted in the provisos above—date of collection ≠ date of origin; simplicity/complexity ≠ anteriority/posteriority, respectively—this cladogram can only offer circumstantial evidence. Indeed, the evolutionarily later placement of **Figure 1Biv** (with its arguably most basic form of M*a*, M*a*[4]) broadly accords with the assertion that simplicity/complexity ≠ anteriority/posteriority.
4. In **Figure 4B**, the exclusive focus on M*c* motivates a restructuring of the cladogram, in that parsimonious relationships of similarity between the alleles of this museme do not always align with parsimonious relationships of similarity between the melodies as a whole (**Figure 4A**). As an example, M*c*[1] is represented as evolutionarily prior to the three other alleles of M*c*, giving **Figures 1Bi,iv** priority; but, in **Figure 4A**, the evolutionarily prior melodies are **Figures 1Bi,iii** (variant). This indeed affords evidence in support of the second claim: that statistical data derived from measuring mutational changes (**Figure 4A**) are epiphenomena of musemic evolution (**Figure 4B**), because M*c* (and indeed any museme) is arguably more meaningful—perceptually-cognitively and evolutionarily—than the larger melody of which it forms a part.
5. In terms of chronology, this second cladogram is (quasi-) anachronistic, in that it ascribes evolutionary (co-)primacy to the "latest" (and also "earliest") of these musemes, M*c*[1]. As specified by the provisos in the third ("chronology") point above, this cladogram does not constitute hard evidence in favor of a phylomemy which runs counter to the collection-date ordering.

This consideration has only scratched the surface of the complex relationships inherent in **Figure 4**, itself only a small case study. For one thing, while these melodies would normally have been performed unaccompanied, their implied harmony (**Figure 1Bv**) may have acted as a selection pressure[30]. Given the tendency for harmonic changes to coincide with points of metrical accentuation—Temperley's "HPR [Harmonic Preference Rule]

[30]Given that unaccompanied melodies in western music normally have clear harmonic implications (a phenomenon arguably most richly developed in the solo violin music of J.S. Bach), the perceptual-cognitive salience of mutations will tend to be evaluated in the light of the silent musemes constituting the underpinning chord progressions. Implied harmony therefore constitutes a selection pressure because it motivates an assessment of the altered conformity of (elements of) a mutant museme with the associated chord vis-à-vis the alignment of its antecedent. In non-western cultures, no such implicative coadaptation exists between melodic and harmonic musemes.

2 (Strong Beat Rule)" (Temperley, 2001, p. 151)—it may be the case that $Ma^{\boxed{1+3}_y}$, with their implied shift to the tonic chord on the second (weak) rather than the third (strong) crotchet beat of the bar (as in $Ma^{\boxed{2+4}_y}$), have either a selective advantage or (paradoxically) a selective *dis*advantage, depending on context[31].

But the overriding issue here is that the dichotomy identified above between real and virtual phylomemies is clearly problematic, for while Savage and Atkinson (2015, p. 167) are laudable in their injunction that statistical-phylomemetic analysis is (only) a stepping stone toward the understanding of "higher-level units of musical structure and meaning", the statistical data—even considered in conjunction with musemic organization—does not always permit the reconstruction of higher-level-unit phylomemies with any real certainty, as is demonstrated by the present study. Perhaps we might simply hypothesize that, in the absence of detailed knowledge of the transmission events under investigation, the cladograms in **Figure 4** *predict* the true temporal ordering of (phrase- or museme-level) events. Thus, we are taking the most parsimonious phylomemy to be the most plausible, and assuming that, when the historical record is obscure, this criterion should be primary when attempting to reconstruct cultural-evolutionary histories.

## 5. CONCLUSION: TWO BROTHERS?

While the lyrics of "The Two Brothers" are decidedly grim, the spirit of this article is optimistic, in that it holds that perceptual-cognitive and statistical models of musical evolution are also brothers (or sisters), and that—unlike the ballad texts—they can go on not to do violence to each other but to grow together and to complement each other, developing to be cooperative adults working for a two-fold common cause: the understanding of cultural evolution as a subset of a wider Darwinian view; and the development of methodologies along the perceptual-cognitive–statistical continuum to investigate its operation.

To return to the two claims underpinning the argument here—(i) that a purely statistical approach based on counting note-edits without consideration of perceptual-cognitive aspects gives an incomplete account of cultural evolution; and (ii) that statistical data derived from measuring mutational changes, while illuminating, are epiphenomena of musemic evolution—we might assert that both have been supported by the (admittedly limited) case study outlined here. That is [apropos claim (i)], Savage's (2017) statistical data on "The Two Brothers" are arguably contextualized, enriched and elucidated by considering the musemic structure of the tune-family music-analytically, music-psychologically and music-phylomemetically; and [apropos claim (ii)] the discussion conducted under the third of these rubrics suggests a strong regulatory role for museme-level (as opposed to note-level) processes.

[31]This context includes the likelihood that, for some listeners, such harmonic-rhythm disruptions might be appealing (and therefore selectively advantageous from the museme's perspective), whereas for other listeners the opposite might be the case.

This case study—a small-scale empirical example of how to pursue a novel methodological strategy—is arguably scalable (by means of more systematic use of computer technology) in ways which would foster perceptual-cognitive–statistical collaboration in research on cultural evolution. The methodology for this, which is essentially a formalization and expansion of what is discussed here, is summarized as follows. As will be clear, many of the relevant technologies already exist and so, as is often the case with advances in research, it is largely a matter of synergistic interconnection for this to become a reality.

1. Music databases need to be utilized. To maximize the big-data approach, sizeable databases in an established music-encoding format should be employed (Selfridge-Field, 1997). The *Humdrum Toolkit*'s (Huron, 2002) **kern format is used for several databases, including the Essen Folksong Collection (Schaffrath, 1995), together with various art-music repertoires, and this format can be translated to other encodings, such as MusicXML (MakeMusic, 2016).
2. Algorithms need to be developed to segment and interrogate the encodings in 1 above in order to locate patterns which are (i) perceptually-cognitively meaningful (using criteria drawn from the music-cognition and music-theory literature); and (ii) replicated in two or more contexts—i.e., patterns which satisfy the necessary conditions for existing as musemes. In addition to Savage's (2017) software, many such algorithms for segmentation and pattern-matching have been developed over recent years, often under the stimulus of the aforementioned MIREX project (Lartillot, 2009; Conklin, 2010; Velardo et al., 2016).
3. The outputs of 2 above need to be processed with phylogenetic software in order to reconstruct hypothetical phylomemies of musemes and the works of which they form part. To accomplish this, greater formalization is needed for the encoding of musical elements and for their incorporation into software designed primarily for (biological) phylogenies. For one thing, a **kern/MusicXML–*Phylip* converter might usefully be developed.
4. Prosopographic analysis (Keats-Rohan, 2007), which is a nascent research methodology in historical musicology, could be extended as a means of contextualizing and assigning probabilities to the outputs of 3 above.

While the four points above seem clear in outline, their connection is likely to prove difficult to implement in practice, given the recalcitrant complexity of music and the intricacy of the programming tasks required. Yet success in this venture offers a rich promise: that of reconstructing how music may have been perceived and transmitted across time and place in various human societies; and therefore of offering synchronic overviews and simulacra of once-vibrant, diachronic musical cultures.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## REFERENCES


Adamic, L., Lento, T., Adar, E., and Ng, P. (2014). *The Evolution of Memes on Facebook*. Available online at: https://www.facebook.com/notes/facebook-data-science/the-evolution-of-memes-on-facebook/10151988334203859?notift=notify_me

Ball, J. A. (1984). Memes as Replicators. *Ethol. Sociobiol.* 5, 145–161.

Berlyne, D. E. (1971). *Aesthetics and Psychobiology*. New York, NY: Century Psychology Series. Appleton-Century-Crofts.

Blackmore, S. J. (1999). *The Meme Machine*. Oxford: Oxford University Press.

Blackmore, S. J. (2000). "The Memes' Eye View," in *Darwinizing Culture: The Status of Memetics as a Science*, ed R. Aunger (Oxford: Oxford University Press), 25–42.

Blackmore, S. J. (2010). *Consciousness: An Introduction,* 2nd Edn. London: Taylor and Francis.

Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms,* 2nd Edn. London; New York, NY: Routledge.

Bronson, B. H. (1959). *The Traditional Tunes of the Child Ballads: With Their Texts, According to the Extant Records of Great Britain and North America*. Princeton, NJ: Princeton University Press.

Brown, S. (2000). "The "Musilanguage" Model of Musical Evolution," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 271–300.

Byros, V. (2009). Towards an "Archaeology" of hearing: schemata and eighteenth-century consciousness. *Musica Humana* 1–2, 235–306.

Calvin, W. H. (1996). *The Cerebral Code: Thinking a Thought in the Mosaics of the Mind*. Cambridge, MA: MIT Press.

Cavalli-Sforza, L. L., and Feldman, M. W. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princeton University Press.

Child, F. J., (ed.). (1904). *The English and Scottish Popular Ballads*. Boston, MA; New York, NY: Houghton Mifflin.

Conklin, D. (2010). Discovery of distinctive patterns in music. *Intell. Data Anal.* 14, 547–554. doi: 10.3233/IDA-2010-0438

Cook, K. M. (2015). Text mining and early music: using lexomics in research. *Early Music* 43, 661–665. doi: 10.1093/em/cav085

Cook, N. (2004). "Computational and comparative musicology," in *Empirical Musicology: Aims, Methods, Prospects*, eds E. Clarke and N. Cook (Oxford: Oxford University Press), 103–126.

Cope, D. (2003). Computer analysis of musical allusions. *Comput. Music J.* 27, 11–28. doi: 10.1162/01489260360613317

Cowdery, J. R. (1984). A fresh look at the concept of tune family. *Ethnomusicology* 28, 495–504.

Darwin, C. (2008). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Oxford: Oxford University Press.

Dawkins, R. (1983a). *The Extended Phenotype: The Long Reach of the Gene*. Oxford: Oxford University Press.

Dawkins, R. (1983b). "Universal Darwinism," in *Evolution from Molecules to Men*, ed D. S. Bendall (Cambridge: Cambridge University Press), 403–425.

Dawkins, R. (1989). *The Selfish Gene*. 2nd Edn. Oxford: Oxford University Press.

Dawkins, R. (2006). *The Blind Watchmaker*. London: Penguin.

Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. London: Penguin.

Dennett, D. C. (2007). *Breaking the Spell: Religion as a Natural Phenomenon*. London: Penguin.

Deutsch, D. (1999). "Grouping mechanisms in music," in *The Psychology of Music*, 2nd Edn., ed D. Deutsch (San Diego, CA: Academic Press), 299–348.

Dronamraju, K. (2010). *Haldane, Mayr, and Beanbag Genetics*. Oxford: Oxford University Press.

Durham, W. H. (1991). *Coevolution: Genes, Culture, and Human Diversity*. Stanford, CA: Stanford University Press.

Eck, R. V., and Dayhoff, M. O. (1966). *Atlas of Protein Sequence and Structure 1966*. Silver Spring, MD: National Biomedical Research Foundation.

Felsenstein, J. (2016). *Phylip v. 3.695*. Retrieved from: http://evolution.genetics.washington.edu/phylip.html

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press.

Gjerdingen, R. O. (1988). *A Classic Turn of Phrase: Music and the Psychology of Convention*. Philadelphia, PA: University of Pennsylvania Press.

Goehr, L. (1992). *The Imaginary Museum of Musical Works: An Essay in the Philosophy of Music*. Oxford: Clarendon Press.

Grigoryan, N. (2011). Science and art: their reciprocal advance. *Herald Russ. Acad. Sci.* 81, 463–467.

Hennig, W. (1999). *Phylogentic Systematics, 3rd Edn*. Urbana, IL: University of Illinois Press.

Hepokoski, J. A., and Darcy, W. (2006). *Elements of Sonata Theory: Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. New York, NY; Oxford: Oxford University Press.

Howe, C. J., and Windram, H. F. (2011). Phylomemetics: evolutionary analysis beyond the gene. *PLoS Biol.* 9:e1001069. doi: 10.1371/journal.pbio.1001069

Huron, D. (2002). Music information processing using the Humdrum Toolkit: concepts, examples, and lessons. *Comput. Music J.* 26, 11–26. doi: 10.1162/014892602760137158

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Husserl, E. (2013). *Zur Phänomenologie des Inneren Zeitbewußtseins*. Hamburg: Felix Meiner.

Jablonka, E., and Lamb, M. J. (2014). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge, MA; London: MIT Press.

Jan, S. B. (2007). *The Memetics of Music: A Neo-Darwinian View of Musical Structure and Culture*. Aldershot: Ashgate.

Jan, S. B. (2010). Memesatz contra Ursatz: memetic perspectives on the aetiology and evolution of musical structure. *Musicae Sci.* 14, 3–50. doi: 10.1177/102986491001400101

Jan, S. B. (2011). Music, memory, and memes in the light of Calvinian neuroscience. *Music Theory Online* 17. Available online at: http://www.mtosmt.org/issues/mto.11.17.2/mto.11.17.2.jan.html

Jan, S. B. (2013). Using galant schemata as evidence for universal Darwinism. *Interdiscipl. Sci. Rev.* 38, 149–168. doi: 10.1179/0308018813Z.00000000042

Jan, S. B. (2014). Similarity continua and criteria in memetic theory and analysis. *J. Music Res. Online* 5. Available online at: http://www.jmro.org.au/index.php/mca2/article/view/125

Jan, S. B. (2015). Memetic perspectives on the evolution of tonal systems. *Interdiscipl. Sci. Rev.* 40, 143–165. doi: 10.1179/0308018815Z.000000000110

Jan, S. B. (2016). A memetic analysis of a phrase by Beethoven: Calvinian perspectives on similarity and lexicon-abstraction. *Psychol. Music* 44, 443–465. doi: 10.1177/0305735615576065

Juhász, Z., and Sipos, J. (2010). A comparative analysis of Eurasian folksong corpora, using self-organising maps. *J. Interdiscipl. Music Stud.* 4, 1–16. doi: 10.4407/jims.2009.11.005

Keats-Rohan, K. S. B. (2007). *Prosopography Approaches and Applications: A Handbook*. Occasional publications of the Oxford Unit for Prosopographical Research. Unit for Prosopographical Research, Linacre College, University of Oxford, Oxford.

Kirby, S. (2013). "Transitions: the evolution of linguistic replicators," in *The Language Phenomenon: Human Communication from Milliseconds to Millennia*, eds P. M. Binder and K. Smith (Berlin; Heidelberg: Springer), 121–138.

Kuper, A. (2000). "If memes are the answer, what is the question?," in *Darwinizing Culture: The Status of Memetics as a Science*, ed R. Aunger (Oxford: Oxford University Press), 175–188.

Lartillot, O. (2009). Taxonomic categorisation of motivic patterns. *Musicae Scientiae, Discussion Forum* 4B, 25–46.

Lerdahl, F. (1992). Cognitive constraints on compositional systems. *Contemp. Music Rev.* 6, 97–121.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Phys. Doklady* 10, 707–710.

Lewontin, R. C. (1970). The units of selection. *Ann. Rev. Ecol. Syst.* 1, 1–18.

Lynch, A. (1998). Units, events and dynamics in memetic evolution. *J. Memetics Evol. Models Inf. Transm.* 2. Available online at: http://jom-emit.cfpm.org/1998/vol2/lynch_a.html

MakeMusic (2016). MusicXML. Available online at: www.musicxml.com

Margulis, E. H., and Beatty, A. P. (2008). Musical style, psychoaesthetics, and prospects for entropy as an analytic tool. *Comput. Music J.* 32, 64–78. doi: 10.1162/comj.2008.32.4.64

Martindale, C. (1986). Aesthetic evolution. *Poetics* 15, 439–473.

McCann, G. (1999). *Morecambe and Wise*. London: Fourth Estate.

McNamara, A. (2011). Can we measure memes? *Front. Evol. Neurosci.* 3, 1–7. doi: 10.3389/fnevo.2011.00001

Merker, B. (2012). "The vocal learning constellation: imitation, ritual culture, encephalization," in *Music, Language, and Human Evolution*, ed N. Bannan (Oxford: Oxford University Press), 215–260.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago, IL: University of Chicago Press.

Meyer, L. B. (1996). *Style and Music: Theory, History, and Ideology*. Chicago, IL: University of Chicago Press.

Mhatre, H., Gorchetchnikov, A., and Grossberg, S. (2012). Grid cell hexagonal patterns formed by fast self-organized learning within entorhinal cortex. *Hippocampus* 22, 320–334. doi: 10.1002/hipo.20901

Miranda, E. R., Kirby, S., and Todd, P. M. (2003). On computational models of the evolution of music: from the origins of musical taste to the emergence of grammars. *Contemp. Music Rev.* 22, 91–110. doi: 10.1080/0749446032000150915

Mithen, S. (2006). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld & Nicolson.

Narmour, E. (1977). *Beyond Schenkerism: The Need for Alternatives in Music Analysis*. Chicago, IL: University of Chicago Press.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago, IL: University of Chicago Press.

Patel, A. D. (2008). *Music, Language, and the Brain*. New York, NY: Oxford University Press.

Pearce, M. T., and Wiggins, G. A. (2012). Auditory expectation: the information dynamics of music perception and cognition. *Top. Cogn. Sci.* 4, 625–652. doi: 10.1111/j.1756-8765.2012.01214.x

Perry, J. (2000). Music, evolution and the ladder of progress. *Music Theory Online* 6. Available online at: http://www.mtosmt.org/issues/mto.00.6.5/mto.00.6.5.perry.html

Podlipniak, P. (2017). The role of the Baldwin effect in the evolution of human musicality. *Front. Neurosci.* 11:542. doi: 10.3389/fnins.2017.00542

Popper, K. (1959). *The Logic of Scientific Discovery*. London: Hutchinson.

Richerson, P. J., Boyd, R., and Henrich, J. (2010). Gene-culture coevolution in the age of genomics. *Proc. Natl. Acad. Sci. U.S.A.* 107(Suppl. 2), 8985–8992. doi: 10.1073/pnas.0914631107

Ridley, M. (2004). *Evolution*, 3rd Edn. Oxford: Blackwell.

Robinson, M. A. (2001). New testament textual criticism: the case for Byzantine priority. *TC* 6. Available online at: http://www.reltech.org/TC/v06/Robinson2001.html

Rose, S., Tuppen, S., and Drosopoulou, L. (2015). Writing a big data history of music. *Early Music* 43, 649–660. doi: 10.1093/em/cav071

Rosen, C. (1988). *Sonata Forms,* 2nd Edn. New York, NY: Norton.

Savage, P. E. (2017). *Measuring the Cultural Evolution of Music: With Case Studies of British-American and Japanese Folk, Art, and Popular Music*. PhD thesis, Tokyo University of the Arts.

Savage, P. E., and Atkinson, Q. D. (2015). "Automatic tune family identification by musical sequence alignment," in *Proceedings of the 16th International Society for Music Information Retrieval Conference* (October 2015) (Malaga), 162–168.

Schaffrath, H. (1995). *The Essen Folksong Collection*. Stanford, CA: Center for Computer Assisted Research in the Humanities.

Schulkin, J. (2013). *Reflections on the Musical Mind: An Evolutionary Perspective*. Princeton, NJ: Princeton University Press.

Selfridge-Field, E. (1997). *Beyond MIDI: The Handbook of Musical Codes*. Cambridge, MA: MIT Press.

Sharma, S., Tim, U. S., Wong, J., Gadia, S., and Sharma, S. (2014). A brief review on leading big data models. *Data Sci. J.* 13, 138–157. doi: 10.2481/dsj.14-041

Shennan, S. (2002). *Genes, Memes, and Human History: Darwinian Archaeology and Cultural Evolution*. London: Thames and Hudson.

Snow, C. P. (1964). *The Two Cultures*. Cambridge: Cambridge University Press.

Snyder, B. (2000). *Music and Memory: An Introduction*. Cambridge, MA: MIT Press.

Tagg, P. (2016). *Kojak: 50 Seconds of Television Music. Towards the Analysis of Affect in Popular Music*. Göteborg University Musicology Department; Mass Media Music Scholars' Press. Available online at: http://www.tagg.org/mmmsp/kojak.html

Taruskin, R. (1995). *Text and Act: Essays on Music and Performance*. New York, NY: Oxford University Press.

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, MA: MIT Press.

Typke, R., Wiering, F., and Veltkamp, R. C. (2007). Transportation distances and human perception of melodic similarity. *Musicae Sci.* 11, 153–181. doi: 10.1177/102986490701100107

Velardo, V. (2016). Recursive ontology: a systemic theory of reality. *Axiomathes* 26, 89–114. doi: 10.1007/s10516-015-9272-0

Velardo, V., and Vallati, M. (2016). "Beyond anthropocentric creativity: a general framework for describing creative systems," in *International Conference on Tools with Artificial Intelligence* (San Jose, CA).

Velardo, V., Vallati, M., and Jan, S. B. (2016). Symbolic melodic similarity: state of the art and future challenges. *Comput. Music J.* 40, 70–83. doi: 10.1162/COMJ_a_00359

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2018 Jan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 April 2018
doi: 10.3389/fpsyg.2018.00381

# Computational Approach to Musical Consonance and Dissonance

*Lluis L. Trulla[1], Nicola Di Stefano[2]* and Alessandro Giuliani[3]*

*[1] Centre de Recerca Puig Rodó, Girona, Spain, [2] Institute of Philosophy of Scientific and Technological Practice and Laboratory of Developmental Neuroscience, Università Campus Bio-Medico di Roma, Rome, Italy, [3] Environment and Health Department, National Institute of Health, Rome, Italy*

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky, Independent Researcher, Los Angeles, CA, United States*

***Reviewed by:***
*Juan G. Roederer, American Association of Retired Persons, United States*
*Susan Elizabeth Rogers, Berklee College of Music, United States*

****Correspondence:***
*Nicola Di Stefano*
*n.distefano@unicampus.it*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *28 August 2017*
***Accepted:*** *08 March 2018*
***Published:*** *04 April 2018*

***Citation:***
*Trulla LL, Di Stefano N and Giuliani A (2018) Computational Approach to Musical Consonance and Dissonance. Front. Psychol. 9:381. doi: 10.3389/fpsyg.2018.00381*

In sixth century BC, Pythagoras discovered the mathematical foundation of musical consonance and dissonance. When auditory frequencies in small-integer ratios are combined, the result is a harmonious perception. In contrast, most frequency combinations result in audible, off-centered by-products labeled "beating" or "roughness;" these are reported by most listeners to sound dissonant. In this paper, we consider second-order beats, a kind of beating recognized as a product of neural processing, and demonstrate that the data-driven approach of Recurrence Quantification Analysis (RQA) allows for the reconstruction of the order in which interval ratios are ranked in music theory and harmony. We take advantage of computer-generated sounds containing all intervals over the span of an octave. To visualize second-order beats, we use a glissando from the unison to the octave. This procedure produces a profile of recurrence values that correspond to subsequent epochs along the original signal. We find that the higher recurrence peaks exactly match the epochs corresponding to just intonation frequency ratios. This result indicates a link between consonance and the dynamical features of the signal. Our findings integrate a new element into the existing theoretical models of consonance, thus providing a computational account of consonance in terms of dynamical systems theory. Finally, as it considers general features of acoustic signals, the present approach demonstrates a universal aspect of consonance and dissonance perception and provides a simple mathematical tool that could serve as a common framework for further neuro-psychological and music theory research.

**Keywords: beating, recurrence quantification analysis, complex systems, non-linear signal analysis methods, Devil's staircase**

## INTRODUCTION

Beating is the sensation that typically occurs when two sounds with similar frequencies mutually interfere, giving rise to a waveform with a rhythmic oscillation in amplitude. Following the fundamental contribution of Helmholtz's treatise, *On the Sensation of Tone* (1954), first published in 1863, contemporary explanations of consonance are grounded in the notions of beating and complex tones—i.e., sounds displaying a broad array of sinusoidal components (harmonics).

Roederer (2008, p. 35) provides an illuminating classification of the effects of superposing two pure tones depending on where in the listener's auditory system the sounds become entangled. The above mentioned beating is labeled by Roederer as "first-order beating," because it is processed mechanically in the cochlear fluid and along the basilar membrane. Evidence of the physiological basis of first-order beating stems from the fact that its effect disappears when sounds are played separately in different ears—i.e., dichotically. Another kind of first-order beating effect is known as combination tones, which are produced by the non-linear interaction of waves in narrow spaces, such as the body of musical instruments or the inner ear. Combination tones can be considered as the product of two sine waves. A common example is the *terzo suono* theorized by Giuseppe Tartini (see Lohri, 2016). If $a$ and $b$ are two frequencies with $a > b$, then the *terzo suono* is a tone at frequency $a–b$ that is discernible only by the listener, because it is produced inside the inner ear rather than being caused by external air vibrations. Combination tones can be heard across the octave at sound pressure levels (SPLs) of 80 dB or higher, and across part of the octave at 50 dB SPL and above.

At 80 dB (or higher) while maintaining the interval around the octave, a distinct beating can be perceived. This disappears when $f_2 = 2f_1$ (where $f_1$ and $f_2$ represent the two frequencies) and reappears as long as the octave becomes mistuned by a factor ε (i.e., $f_2 = 2f_1 + \varepsilon$). The beating frequency turns out to be ε (Plomp and Levelt, 1965). Beating "is created by the relatively quick changes produced by modulation frequencies in the region between about 15 to 300 Hz" (Fastl and Zwicker, 2006, p. 257). Unlike first-order beats, the beating persists when tones are fed dichotically, implying that, in this case, beat perception is the result of neural processing. Hence, they are defined as "second-order beats" (Roederer, 2008). Second-order beating shows a modulation in the vibration pattern, i.e., a periodic change in phase difference between the two sounds that form the interval (Roederer, 2008, p. 49), although no amplitude modulation is present. Second-order beats are also called "beats of mistuned consonances" because they are audible when pure tones are superposed to form a fifth (Plomp, 1976). In fact, whereas the vibration pattern of a correctly tuned fifth ($f_2 = 3/2\ f_1$) or fourth ($f_2 = 4/3\ f_1$) is static, the mistuned cases $f_2 = 3/2\ f_1 + \varepsilon$ and $f_2 = 4/3\ f_1 + \varepsilon$ cause the vibration pattern to change periodically in form, but not in amplitude. From the octave to the fifth and to the fourth, the second-order beats become faster (beating frequency being ε for the octave, 2ε for the fifth, and 3ε for the fourth) as the vibration pattern grows in complexity (see **Figure 1**).

Their neural origin makes second-order beats an excellent phenomenon for investigating the link between the mathematical description of the signals and their neural processing, and consequently allows us to shed light on their perceived "pleasantness." To achieve a consistent picture of second-order beats, it is fundamental to overcome the frequency–time space representation trade-off and the related problem of non-stationary signal characteristics.

Graphic representations of sound typically plot the course of amplitude over time or report the relative amplitudes of the different frequencies computed by the Fourier Transform. Thus, there is no mention of time in the latter, and no mention of frequency in the former. However, in the actual hearing process, time and frequency are strictly intermingled, because specific frequencies are processed at specific moments. This fact suggests that we should focus on the simultaneous analysis of time/frequency dimensions (Roads, 2001).

To determine the frequency of an oscillatory phenomenon, we must count the number $n$ of vibrations that occur within

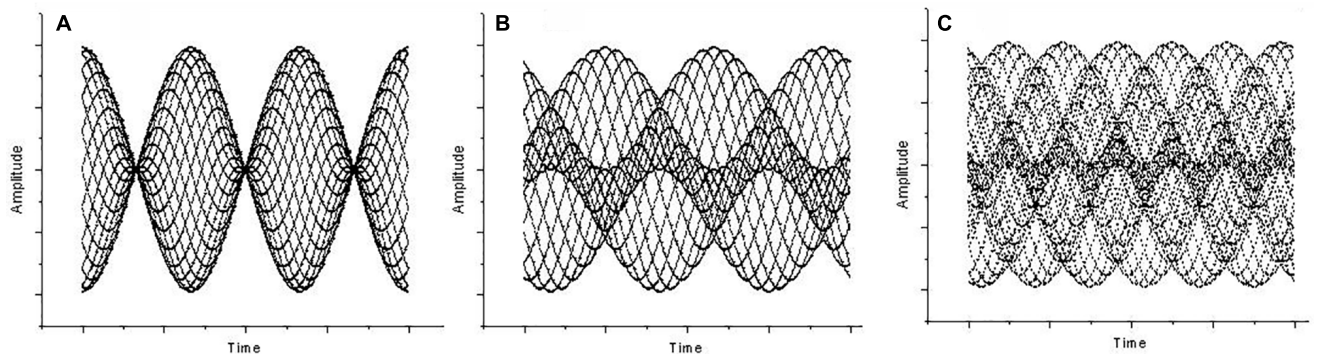



**FIGURE 1 |** Amplitude (y axis) against time (x axis) for **(A)** Mistuned unison, a case of first-order beating (the interval ratio is 400/403 Hz, ε = 3, 3 beats). **(B)** Mistuned octave at 803/400 Hz (ε = 3, 3 beats), a case of second-order beating. **(C)** Mistuned fifth at 603/400 Hz (ε = 3, the beats are 2ε), also a case of second-order beating.

a set time interval $\Lambda t$. As $n$ is an integer, the minimum error in measuring the frequency is one, thus generating a kind of uncertainty principle in the form $\Lambda f \geq 1/\Lambda t$. Increasing the precision of the frequency reclaims a wider window in which to count the time, thus increasing the indetermination of the instant in which the specific frequency occurs.

It is possible to neglect the explicit consideration of time and visualize tone relationships within the octave by computing the ratio of two simultaneous frequencies and then plotting the interval ratio against the amplitude. This is achieved by forming a linear combination of two pure tone waves, a glissando from the unison ($f_1$) to the octave ($2f_1$) and a firm wave at frequency $f_1$. Similar stimuli were previously adopted by Helmholtz (1954) and Kameoka and Kuriyagawa (1969a,b). More recently, Piana (2007) provided a phenomenological explanation of consonance and dissonance when moving from the glissando and ruling out intervals and harmonics. In the following, we propose a numerical approach to Helmholtz's glissando. Note that this approach maintains the time dimension in terms of the determined sequence of interactions between glissando and the fixed frequency. Focusing on these interactions allows us to overcome the trade-off in the frequency–time representation. For this purpose, we base our approach on the concept of recurrence, a simpler and more fundamental property of the signals with respect to the oscillation frequency (Eckmann et al., 1987; Marwan et al., 2007). The degree of recurrence of a series is estimated by the number of times a signal comes back to an already visited state (see section "Materials and Methods"), and can be computed by the application of recurrence quantification analysis (RQA) (Marwan et al., 2007). Estimating the recurrence rate avoids any stationarity assumption, as the estimate is obtained by a "computation window" sliding along the signal; the result is a profile of recurrence values relative to subsequent epochs along the original signal. This provides a model-free, discrete, and local estimation of the recurrent properties of the series, enabling a quantitative description of second-order beats. The recurrence peaks exactly match the values of the interval ratios corresponding to just intonation and are proportional to the order of consonance of the intervals, thus providing a link between consonance and the dynamical features of the signal.

# MATERIALS AND METHODS

## Recurrence Quantification Analysis

The original idea of describing non-stationary signals (which are not amenable to classical Fourier analysis) by means of recurrence dates back to the work of Ruelle's group (Eckmann et al., 1987). The authors introduced recurrence analysis as a purely graphical technique in the form of recurrence plots (RP). Webber and Zbilut (1994) then converted the RP approach into a quantitative technique (RQA) by defining some non-linear descriptors of the RP. RQA has been adopted for the assessment of time series structures in fields ranging from molecular dynamics to physiology and text analysis (Manetti et al., 1999; Orsucci et al., 2006; Marwan et al., 2007). In the field of music research, RQA has been successfully applied to song recognition (Serra et al., 2009) and in the definition of an objective basis of consonance of pure tones (Trulla et al., 2005). In general, this non-linear technique is especially useful for quantifying transient behavior far from the equilibrium (Trulla et al., 1996).

RQA builds upon the computation of a distance matrix between the rows (epochs) of the embedding matrix of the signal of interest, with the lag defined by the method of the first minimum of Mutual Information (Kennel et al., 1992). Given a scalar time series {x(i) = 1; 2; 3;. . .}, an embedding procedure generates a vector Xi = (x(i); x(i+$L$);. . .; x(i+($m$-1)$L$)), where $m$ is the embedding dimension and $L$ is the lag. {Xi = 1; 2; 3;. . .; N} then represents the multi-dimensional process of the time series (signal) as a trajectory in $m$-dimensional space. RPs are symmetrical N × N matrices in which a point is placed at (i; j) whenever a point $X_i$ on the trajectory is close to another point $X_j$. The relative closeness between $X_i$ and $X_j$ is estimated by the Euclidian distance between these two vectors. If the distance falls below a threshold radius ($r$), the two vectors (epochs, windows) are considered to be recurrent, and this is graphically indicated by a dot. The value of $r$ is usually set to 5–10% of the average pairwise distances between epochs. Therefore, RPs correspond to the symmetrical distance matrix between the epochs (rows of the embedding matrix) of the signal transformed into a binary 0/1 matrix by the action of a threshold.

As an example, consider a time series $A$ made up of 10 consecutive values: 7, 8, 10, 15, 6, 7, 9, 11, 10, 8. To observe the recurrence structure of the series at the level of subsequent epochs of length 3, we transform $A$ into the embedding matrix $A$E:

| t0 | t+1 | t+2 | epochs |
|---|---|---|---|
| 7 | 8 | 10 | ep1 |
| 8 | 10 | 15 | ep2 |
| 10 | 15 | 6 | ep3 |
| 15 | 6 | 7 | ep4 |
| 6 | 7 | 9 | ep5 |
| 7 | 9 | 11 | ep6 |
| 9 | 11 | 10 | ep7 |
| 11 | 10 | 8 | ep8 |

Thus, the original series has been projected into a three-dimensional space in which the variables (columns) are the time-lagged original series and the statistical units (rows) are the overlapping epochs. The second step is to compute the Euclidean distances between the epochs. This generates the following distance matrix AD:

| ep1 | ep2 | ep3 | ep4 | ep5 | ep6 | ep7 | ep8 | **TIME** |
|---|---|---|---|---|---|---|---|---|
| 0 | | | | | | | | ep1 |
| 5.477226 | 0 | | | | | | | ep2 |
| 8.602325 | 10.48809 | 0 | | | | | | ep3 |
| 8.774964 | 11.35782 | 10.34408 | 0 | | | | | ep4 |
| **1.732051** | 7 | 9.433981 | 9.273618 | 0 | | | | ep5 |
| **1.41421** | 4.242641 | 8.3666 | 9.433981 | 3 | 0 | | | ep6 |
| 3.605551 | 5.196152 | 5.744563 | 8.3666 | 5.09902 | 3 | 0 | | ep7 |
| 4.898979 | 7.615773 | 5.477226 | 5.744563 | 5.91608 | 5.09902 | 3 | 0 | ep8 |

As the AD elements correspond to the Euclidean distances between corresponding epochs, the diagonal values are 0, and the symmetric character of the distances implies the matrix can be written in lower-triangular form.

We now specify that two epochs are recurrent if their distance is less than 95% of all the between-epoch distances. The average value of the below-diagonal elements of AD is 6.48, and their standard deviation is 2.74. Thus, it is estimated that 95% of distances are greater than 1.74. This implies we have only two recurrences, corresponding to the epoch1–epoch5 and epoch1–epoch6 couples (bolded in the table).

Therefore, example series A has a recurrence rate of 0.071 (two recurrences out of 28 distinct distances) or, equivalently, a recurrence percentage equal to 7.1. The AD matrix corresponds to an RP with only two dots, at coordinates (1, 5) and (1, 6). Note that the recurrences can be identified without the need for any frequency estimation, thus resembling the hearing process that receives sounds as they occur in time.

To provide a quantitative measure of the recurrence, numerical RP descriptors were developed (Marwan et al., 2007). We now consider the proportion of recurrent points (dots) in a plot, called the recurrence. Going back to the music domain, **Figure 2** reports the data relative to **Figure 1** as RPs.

## Software

Files were generated using the sound editor Cool Edit Pro and saved in ASCII format before being fed to the Visual Recurrence Analysis (VRA) software. For the plots in **Figure 1**, we loaded a stereo file of 8000 samples/s to the audio editor, and sent a fixed pure tone of 400 Hz lasting 6 s through the left channel and a fixed pure tone of 403 Hz (**Figure 1A**), 803 Hz (**Figure 1B**), or 603 Hz (**Figure 1C**) for 6 s through the right channel. The sample type was then converted from stereo to mono. **Figure 3** was generated by loading a stereo file of 8000 samples/s to the audio editor, and sending a linearly increasing sound from 360 to 840 Hz lasting 6 s to the left channel and a fixed pure tone of 400 Hz lasting 6 s to the right channel. Finally, the sample type was again converted from stereo to mono.

**Figure 2** shows RPs for the data in **Figure 1**, i.e., 1 s (8000 points) of a mistuned unison, octave, and fifth. The plots were generated by calculating the global recurrence using RQA, as there is no change along the sample. The recurrence of the data shown in **Figure 3** was calculated using a windowing version of an RP, whereby the recurrence is calculated repeatedly for a window that is continuously shifted along the whole sample. Among the RQA parameters, we chose the simplest one, Percent Recurrence, a descriptor that sets the percentage of recurrent points with respect to the non-trivial maximum [equal to (N × (N−1))/2

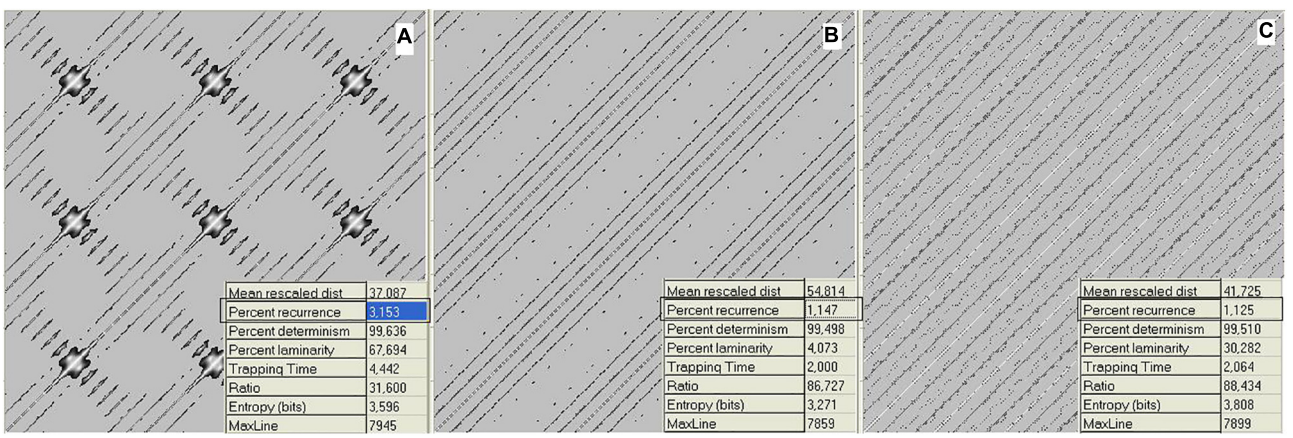



**FIGURE 2 |** Recurrence plots (RPs) of waveforms for mistuned unison **(A)**, mistuned octave **(B)**, and mistuned fifth **(C)**. Calculations were performed on the data in **Figure 1**. Recurrence algorithm generates several descriptors (inset of the figures) of the recurrence distribution. Here, we consider the recurrence parameter. The axis refers to the discrete timing of the signal. RPs are graphical representations of a between-epochs distance matrix (see section "Materials and Methods"). The main diagonal line refers to the coincidence in time, while increasing distances (along both directions) correspond to the recurrences found at increasing delays. The individual dots denote the epoch pairs that have a distance value below the threshold and are thus considered to be recurrent.

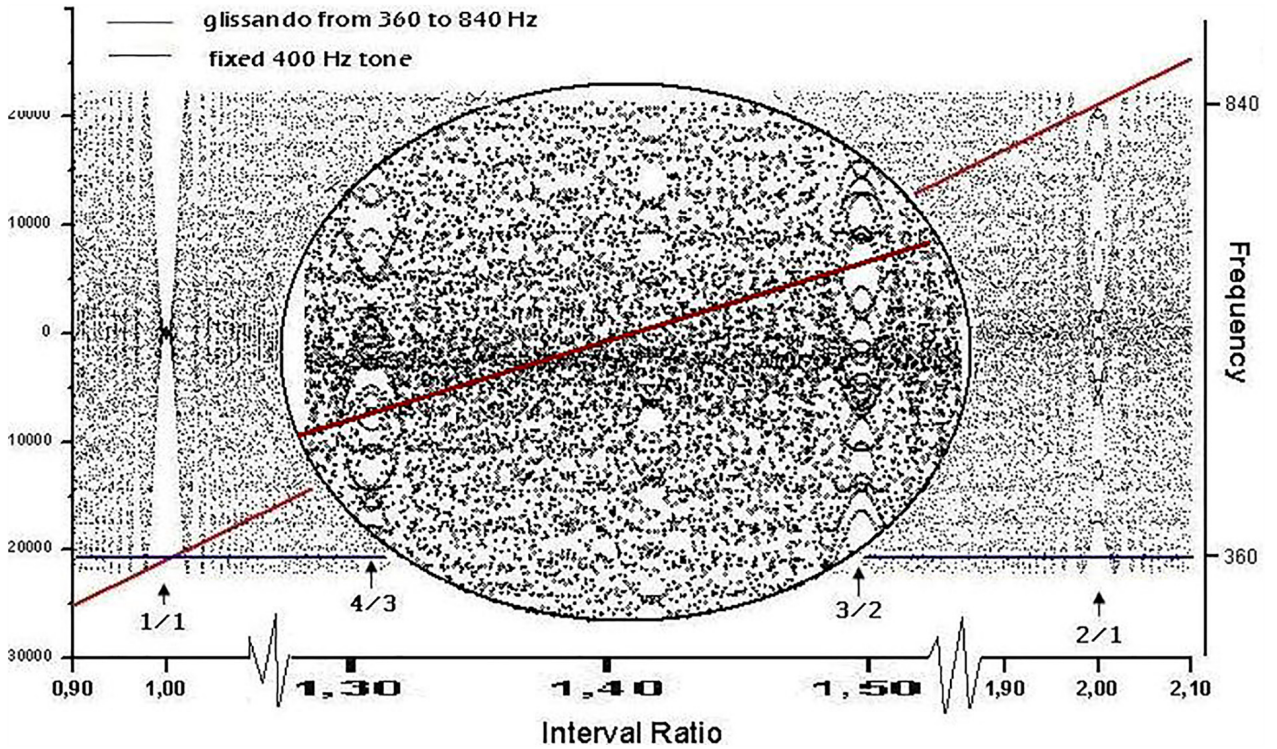



**FIGURE 3 |** Waveform resulting from linearly adding the amplitudes of two sinusoidal signals: a glissando from 360 to 840 Hz (represented by the diagonal line) and a constant frequency of 400 Hz (line parallel to the x axis). The left y axis shows the amplitude of the waveform and the right y axis is the frequency of the diagonal and plain lines. The x axis shows the time for the glissando to go from 360 to 840 Hz, and therefore contains the full collection of intervals between 360/400 and 840/400. The waveform exhibits a rich texture, as the zoomed inset shows, where the intervals of fourth (4/3) and fifth (3/2) are marked. The discrete character of the signal is the cause of the dot-like nature of the graph. The y axis has both negative and positive numbers depending upon the peak/valley alternation of the combination (where anti-phase destructive interactions correspond to 0).

for an N-point series]. The window for recurrence analysis was 480 points long and the shift was 48 points. The embedding dimension was 5 and the delay was 3 points.

MATLAB programs were obtained from http://sethares.engr.wisc.edu/consemi.html for Sethares' dissonance curve and http://courses.theophys.kth.se/5A1352/mfiles/devils.m for the theoretical Devil's staircase (see Discussion).

## RESULTS

A non-stationary signal exploring all interval combinations within the octave can be generated by merging the course of two sounds into a single waveform. The first sound is set at constant frequency $f_1$ for the full duration of the course, while the second follows an ascending glissando from $f_1$ to $f_2 = 2f_1$. **Figure 3** shows an instance of the above procedure.

The most conspicuous singularity (recurrence peaks, see below) in the graph occurs when lines cross themselves, i.e., when $f_2 = f_1$ (unison, interval ratio of 1:1). A second relevant case occurs at the interval ratio of 2:1, which corresponds to the octave. Less evident events occur at 3:2 (fifth) and 4:3 (fourth), as can be seen in the zoomed inset in **Figure 3**. Singularities in the waveform are thus localized where the frequency ratios are expressed by lower integers and with an apparent amplitude (or degree of singularity) matching the accepted ranking of consonance. In our representation, second-order beats appear as a zone of relative calm centered in rational numbers, surrounded by the tempestuous region of irrationals that Roederer (2008) called "beat holes."

Following the numerical solution of Helmholtz's glissando, we explore the glissando/constant frequency signal through an RQA windowing procedure called Recurrence Quantification of Epochs (RQE). RQE performs a scansion of the whole signal by sequentially selecting small windows—specifically episodes of 480 points—in which the RQA algorithm (with the consequent computation of recurrence rate for each episode) is applied. The subsequent windows are shifted by 48 points and the process is repeated throughout the entire file. For each iteration, we retain both the recurrence value and the interval ratio in which this value occurs, calculated as the mean of the interval ratios in the window. **Figure 4** represents the degree of recurrence along the continuum of interval ratios within the octave.

Emergent features of the glissando are evident in **Figure 4**. Firstly, the higher peaks exactly correspond to the places of just intonation (see Trulla et al., 2005), thus establishing a link between pleasantness and the dynamical features (i.e.,

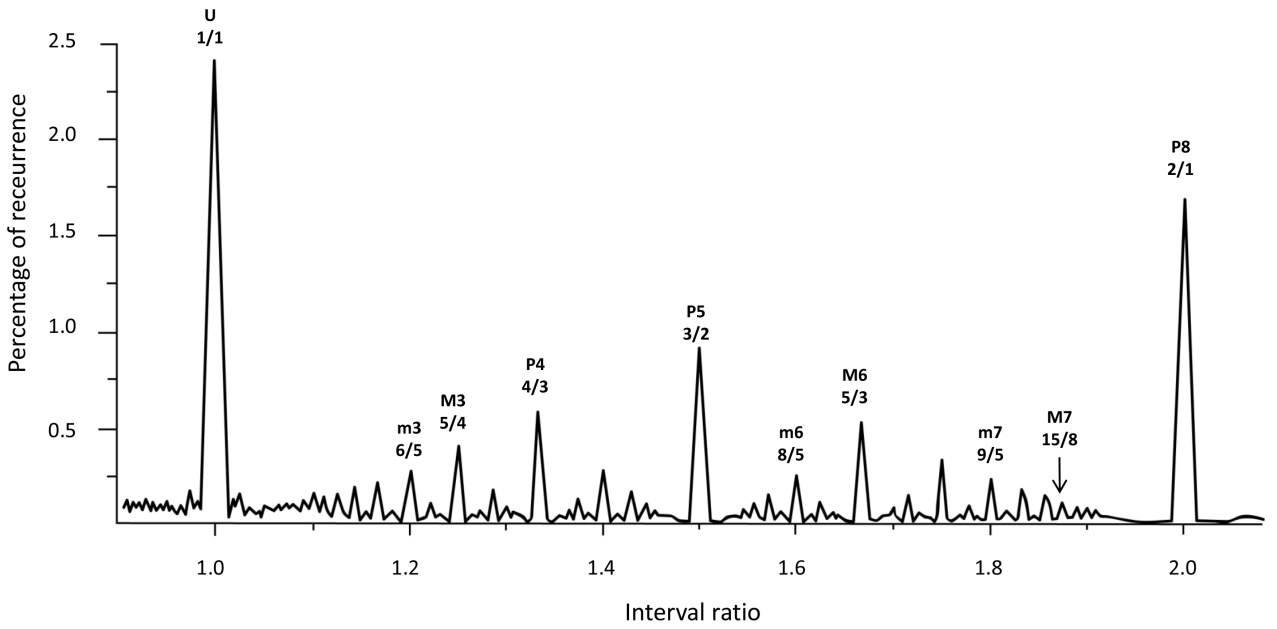



**FIGURE 4 |** Recurrence analysis of the waveform resulting from linearly adding the amplitudes of sinusoidal signals covering the intervals forming the octave. The x axis is the interval ratio and the y axis gives the percentage of recurrence. Each point in this graph is the result of a single recurrence analysis (like those shown in **Figure 2**), from which we obtain the percentage recurrence. In this case, the RQE algorithm performs recurrence analysis over a window of 480 points, retains the percentage recurrence, slides the 480 point window some 48 points, performs recurrence analysis again, and so on until it exhausts the file. The peaks in the graph are labeled with names and rational numbers according to their position along the x axis interval ratio continuum (see **Table 1** for complete interval list and abbreviations).

recurrence) of the signal. As expected from the numerical model, all peaks correspond to rational numbers. Secondly, it is worth noting the symmetry of the peaks around the perfect fifth. Moreover, the correlation between the extent of recurrence and the rank order of consonance derived from the literature evidences the link between the present model based on signal analysis and results from psychological approaches (see Schwartz et al., 2003, i.e., U > P8 > P5 > P4 > M6 > M3 > m3 > m6 > m7 > M7, in decreasing order of consonance; see **Table 1**).

In summary, RQA allows us to establish a natural link between the signal properties and the consonance judgment of the listeners without any *a priori* hypothesis or frequency estimation. The reasons why integer numbers play such an important role in harmony has recently been addressed in the literature, with many different recipes presented for calculating the simplicity of the intervals. We use the consonance index provided by Frova (1999) to demonstrate the close relationship between the proposed recurrence index and the bare numerical characteristics of the intervals. If $m/n$ is the rational number in its lowest terms, Frova's index is $(m+n)/(m \times n)$ (Frova, 1999, p. 178). **Figure 5** illustrates the correlation of this index with the notion of simplicity (i.e., degree of recurrence).

Whereas Frova's index is derived from the energy of the partials forming a complex sound, the percentage recurrence is a purely bottom–up phenomenological descriptor of a pure tone signal, relating recurrence (and consonance) to secondary beating and thus providing a natural (albeit roughly phenomenological) link between the signal properties and neural processing.

Note that the computation of recurrences gives very similar results with respect to models based on primary beating, such as the Plomp and Levelt model reported in **Figure 6**.

## DISCUSSION

In this paragraph, we relate the self-similar appearance of the recurrence graph in **Figure 4** to the mathematical fractal structures generated by physical processes. **Figure 7** shows the empirical cumulative recurrence distribution (obtained by adding consecutive points) and a formal *Devil's staircase* in the [1, 2] interval: the similarities between the two graphs are remarkable.

The *Devil's staircase* pattern is a fingerprint of dynamical systems characterized by the mode-locking phenomenon (Schroeder, 1990, p. 171), which is crucially important in both music generation and perception. In the $17^{th}$ century, Christian Huygens studied mode-locking and discovered the phenomenon of resonance. He noticed that, after a time, the pendulums of two clocks fixed on the same mounting swung synchronously. The synchronization of two coupled oscillators starting from (slightly) different frequencies is called resonance. A more general case of resonant behavior appears when a specific constant frequency is periodically driven by an external power to oscillate at a different frequency; the so-called Devil's staircase

pattern refers to the behavior of forced *quasilinear oscillators*. In the glissando, the constant frequency is the intrinsic frequency and the glissando the external periodic force. Every plateau in the Devil's staircase relates to a particular phase-locked solution (stable state), and its relative width forms a hierarchy that follows the explained propriety of rational numbers. The mathematical model for this case is the circle sine map (McCauley, 1994).

$$\theta n + 1 = \theta n + \frac{p}{q} + (\frac{k}{2\pi}) \sin(2\pi\theta n)$$

where $k$ is a coupling strength parameter that controls the degree of non-linearity. Without coupling ($k$ = 0), the behavior of the system is expressed by the ratio $p/q$ (often called **Ω**, the bare winding number). When $k > 0$, the system locks into rational frequency ratios, preferably with small denominators. In this case, the long-term description of the system corresponds to $w$, the dressed winding number. For the critical value $k$ = 1, the infinite number of locked frequency intervals corresponding to all the rational numbers between 0 and 1 cover the entire **Ω** range.

In our terms, **Ω** is the cumulative recurrence and $w$ is the interval ratio. In other words, the system is locked at any rational number—indicated as the interval ratio—but the width or extent of the lock comes from the cumulative recurrence. Thereby, most relevant consonances have extended areas around the lowest rationales—like the unison or octave—and a strong attraction exists toward these exact ratios. This is perfectly sound in terms of music theory.

The above considerations can be summarized in three main points:

(1) A purely empirical, data-driven analysis (RQA) has highlighted a fundamental property of signals (recurrence distribution) that matches the mathematical (number theory) and physical (mode-locking) theoretical background.
(2) The empirical results are consistent with both a theory-driven "simplicity index" (Frova's index) and with the order that music intervals are ranked in harmony.
(3) The focus on signal properties (second-order beatings) allows us to consider our results as a basis for modeling consonance and dissonance perception by combining data from both computational and cognitive models, e.g., based on artificial neural networks and Hebbian neuroplasticity (Pankovski and Pankovska, 2017).

Numerous studies have confirmed the adequacy of concepts from non-linear dynamics for music perception and construction (e.g., Cartwright et al., 2001, 2002, 2010), and for the study of synchronization among sound sources (Abel et al., 2009). Additionally, neuroscientific research has adopted non-linear dynamical models to describe phase-locked neural populations

**TABLE 1 |** Rank order of consonances and their degree of recurrence.

| Recurrence | Interval ratio | Label | Rational | Name |
|---|---|---|---|---|
| 100,0 | **0.9999** | **U** | **1/1** | **Unison** |
| 89,1 | **2.0006** | **P8** | **2/1** | **Octave** |
| 45,2 | **1.5003** | **P5** | **3/2** | **Perfect fifth** |
| 30,6 | **1.3335** | **P4** | **4/3** | **Perfect fourth** |
| 29,6 | **1.6671** | **M6** | **5/3** | **Major sixth** |
| 23,4 | **1.2495** | **M3** | **5/4** | **Major third** |
| 19,9 | 1.7499 | H7 | 7/4 | Harmonic seventh |
| 18,5 | **1.2003** | **m3** | **6/5** | **Minor third** |
| 16,3 | 1.4007 | | 7/5 | Septimal |
| 15,4 | **1.5999** | **m6** | **8/5** | **Minor sixth** |
| 15,1 | **1.8003** | **m7** | **9/5** | **Just minor seventh** |
| 14,2 | 1.1667 | | 7/6 | Septimal minor third |
| 11,9 | 1.2855 | | 9/7 | Septimal major third |
| 11,7 | 1.8339 | | 11/6 | Undecimal neutral seventh |
| 11,5 | 1.1427 | | 8/7 | Septimal whole tone |
| 10,1 | 1.4283 | | 10/7 | Euler's tritone |
| 9,8 | **1.1247** | **Mt** | **9/8** | **Major whole tone** |
| 9,7 | 1.7139 | | 12/7 | Septimal major sixth |
| 9,4 | 1.5711 | | 11/7 | Undecimal augmented fifth |
| 9,3 | **1.1115** | **mt** | **10/9** | **Minor whole tone** |
| 9,1 | 1.8567 | | 15/8 | Classic major seventh |
| 9,1 | 1.2219 | | 11/9 | Undecimal neutral third |
| 8,7 | 1.1006 | | 11/10 | 4/5 tone |
| 8,6 | 1.3755 | | 11/8 | Undecimal semi-augmented fourth |
| 7,8 | **1.8747** | **M7** | **15/8** | **Classic major seventh** |
| 7,7 | 1.2999 | | 13/10 | Tridecimal semi-diminished fourth |
| 7,6 | 1.6251 | | 13/8 | Tridecimal neutral sixth |
| 7,1 | 1.0911 | | 12/11 | 3/4 tone |
| 6,8 | 1.8891 | | 17/9 | Septendecimal minor third |
| 6,8 | 1.1823 | | 13/11 | Tridecimal minor third |
| 6,7 | **1.4451** | **D5** | **13/9** | **Tridecimal diminished fifth** |

*Bolded values are intervals most used in Western harmony.*

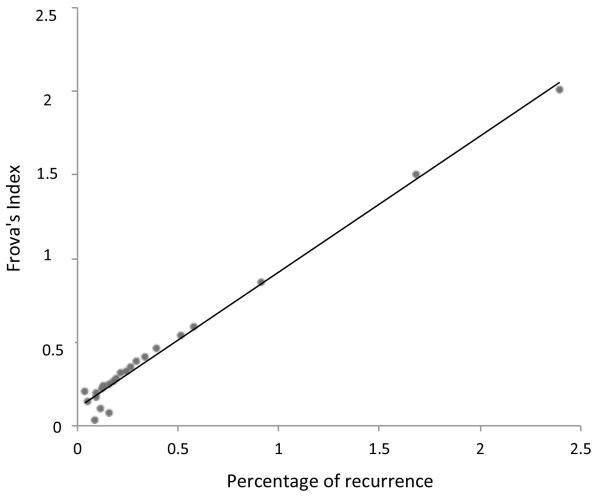



**FIGURE 5 |** Linear relationship between the degree of recurrence (**Figure 4**) and Frova's index of consonance (Frova, 1999). Note the almost perfect overlap between the *a posteriori* statistics of actual signals (i.e., recurrence) and the theoretically motivated *a priori* consonance index (i.e., Frova's index).

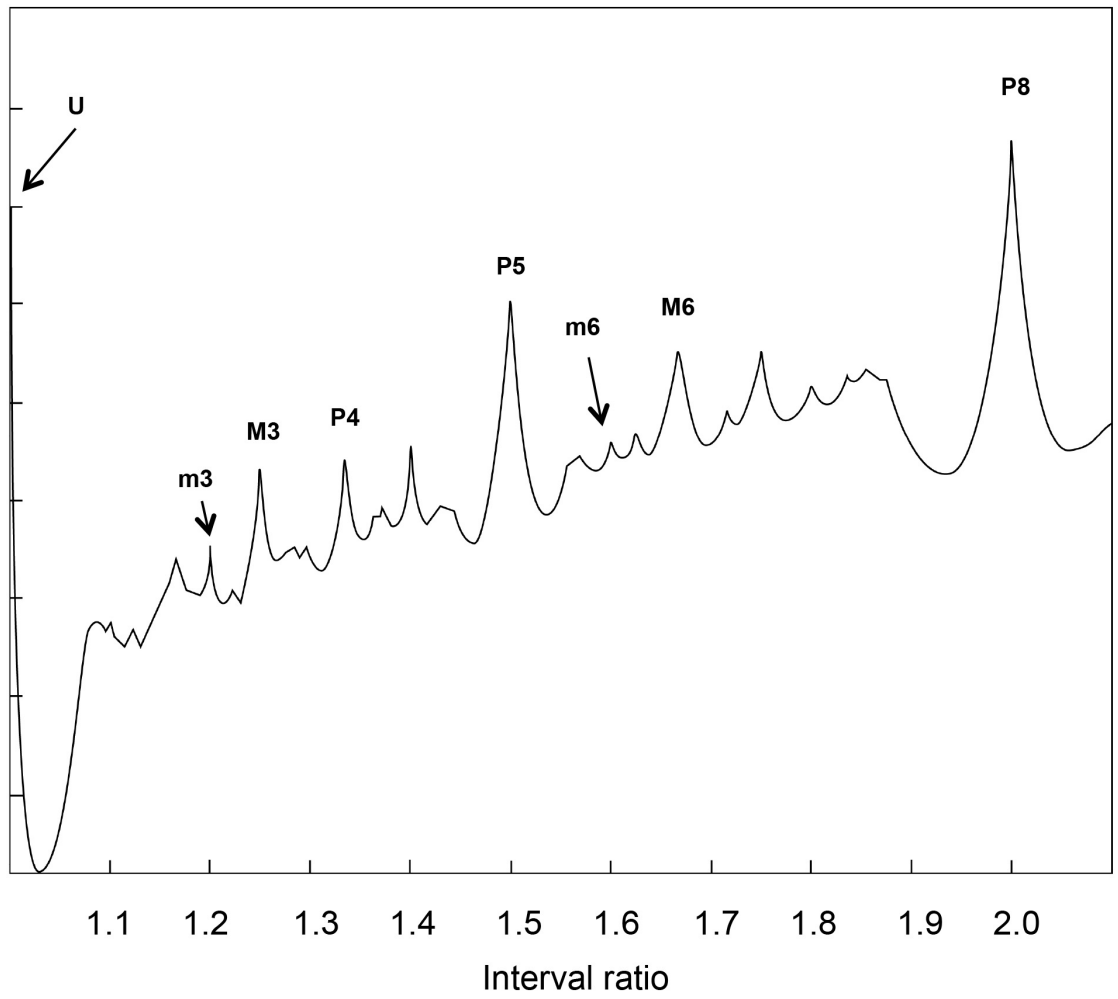



**FIGURE 6 |** Dissonance curve derived from a synthetic sound with 15 harmonics following a natural series. This graph comes from an algorithm ideated by Plomp and Levelt (1965). The plot is shown upside down for ease of comparison with **Figure 4**. The resemblance between the peaks of **Figure 4** and this figure allows for a straightforward interpretation of recurrence results in terms of consonance/dissonance.

(Bidelman and Krishnan, 2009; Large and Almonte, 2012) and build *in silico* neuronal models (Lots and Stone, 2008).

Taken together, our work and previous results support the idea that the production and perception of sound are intimately linked, the perceived pleasantness of intervals being an intrinsic property of the signal (in terms of the degree of recurrence), and not only a secondary effect of the signal on the listener. In turn, this allows us to speculate on the auditory system. Second-order beats have been attributed to the central auditory nervous system, and neuronal webs are known to support phase-locking, as in the mammalian auditory system, in which neural activity in areas including the cochlear nucleus, inferior colliculus, and primary auditory cortex is phase-locked to the stimulus waveform (Large and Tretakis, 2005). The mode-locking model was proposed by Lots and Stone (2008) as the basis for musical consonance, leading to the development of a dynamical system model based on stylized neural oscillators producing both synchronization and mode-locking. These results support the idea that both parts of the communication system (the sender and the receiver of sounds) are similarly "wired." Bidelman and Heinz (2011) applied a waveform to a computational model of the acoustic nerve and, after deriving the autocorrelation function for the nerve fibers, generated the pitch salience profile for the different intervals, giving rise to a distribution that could be superimposed onto the recurrence rate (**Figure 4**). Using an artificial neural network model, Pankovski and Pankovska (2017) recently demonstrated that a specific auditory spectral distribution caused by non-linearities and Hebbian neuroplasticity are sufficient phenomena for a system to generate the consonance pattern.

In line with the literature on music perception (Benade, 1973; Roederer, 2008), we believe that the link between music generation and perception could rely on the fact that the vibrating elements of musical instruments undergo mode-locking into stationary complex vibration patterns. In turn, these can be recognized as the "best fit" to a harmonic template (resident in a properly wired neural circuit). Though this explanation

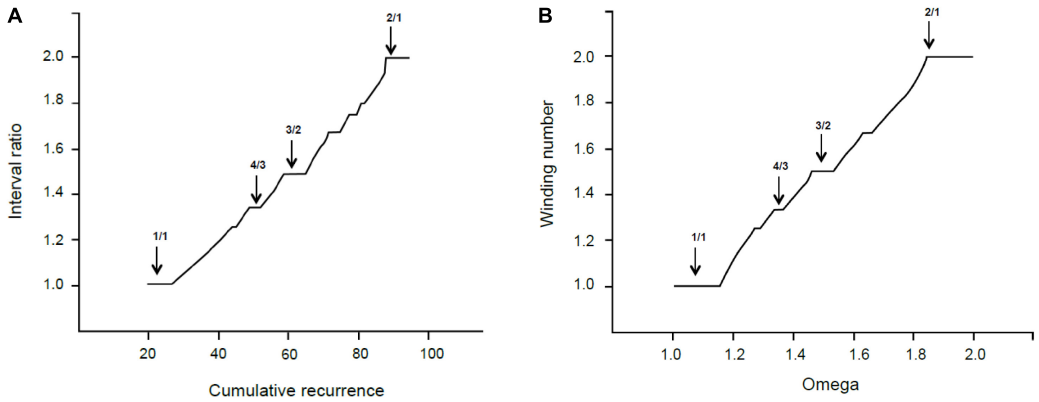



**FIGURE 7 | (A)** Interval ratio vs. cumulative recurrence. **(B)** Theoretical Devil's staircase from sine map with $k = 1$.

stems from empirical correlations, we are convinced that the simplicity and versatility of the RQA approach could pave the way for neuro-psychological studies with the great advantage of considering the acoustic signal and the perceiver from the same mathematical perspective.

The origins of the distinction between consonance and dissonance have been hotly debated in recent years. As the phenomenon of consonance represents a key element of Western music theory, this has mainly been investigated in terms of Western science (i.e., mathematics, physics, psychoacoustics, and neuroscience). For this reason, Parncutt and Hair (2011) called for studies on the use of consonance and dissonance in non-Western cultures to be conducted in terms of local indigenous musicians, rather than in terms of Western science. In this direction, a relevant study published in *Nature* by McDermott et al. (2016) compares the harmonic preferences of people who have different degrees of exposure to Western music. An indigenous population from Bolivia (the Tsimané) was assumed to have no exposure to Western music, and their preferences were compared with groups of city residents in Bolivia and the United States with different degrees of exposure to Western music. The results show that the subjective preferences of Tsimané participants differ from those of the comparison groups; in particular, they failed to rate consonance as being more pleasant than dissonance. The authors state that, as the Tsimané are able to hear the acoustic distinctions associated with consonance and dissonance, the lack of a measurable preference for consonance appears to reflect difference in their aesthetic response to this contrast (McDermott et al., 2016, p. 549). Correctly, they state that the observed cross-cultural variation suggests that consonance preferences are unlikely to be innate, and so preference is probably acquired. However, the fact that the preference for consonance co-varies with presumptive exposure to Western culture is not sufficient to conclude that consonance perception is not biologically determined. Though preferences vary with cultures, the discrimination of consonance is a prerequisite for preference and has a biological basis, as supported by a large number of neurobiological studies (Koelsch and Mulder, 2002; Koelsch et al., 2005; Minati et al., 2008; Perani et al., 2010; Park et al., 2011; Wang, 2013). Investigating whether consonance perception is biologically determined or shaped by culture is likely to be misleading, as it conceives enculturation as a non-biologically constrained process. Harmonic intervals are a consequence of the entrainment of the nervous system with the sound excitation. This forms a universal biological foundation under any musical culture, determining the distinction between acoustic consonance and dissonance and leaving it to each culture to determine exactly how to employ these acoustic distinctions. However, the existence of different musical cultures and systems does not imply the lack of a shared natural/biological basis for music production. The interaction between nature and culture is much more complex, and cross-cultural variations in musical systems only show that biology does not rigidly determine music aesthetics. Similar considerations have led to a more adequate definition of music as a "biocultural phenomenon" (Cross, 2003).

## CONCLUSION

The main contribution of this paper stems from the numerical solution of Helmholtz's glissando. Though the standard modern theory of consonance is based on first-order beating, we have shown that similar results can be obtained starting from second-order beats. The recent interest in second-order beating has been fruitful for models of pitch recognition or neural circuitry (see Roederer, 2008), but not for theories on consonance.

Scholars have started to consider music from the perspective of dynamical systems, both in neurobiological and physical terms, showing that mode-locking models can explain how the nervous system manages sound and is engaged in the ranking of consonances. The resemblance between the formal Devil's staircase model and the cumulative recurrence distribution strengthens this idea.

From a methodological perspective, the main contribution of this work is to provide neuroscience scholars with an extremely simple and model-free tool (RQA) that approaches the acoustic signal and the listener's perception system with the same mathematical method. Different RQA applications have

been reported in research on otoacoustic emission (see, for example, Zimatore et al., 2002, 2003). We are therefore confident that the use of a simple statistical approach will foster interactions between music theory and neuro-psychological approaches.

Finally, our results support the idea of natural roots of consonance perception, and are thus in line with several studies published in recent years (see, for example, Wang, 2013; Bowling and Purves, 2015; Nikolsky, 2015; Foo et al., 2016; González-García et al., 2016; Di Stefano et al., 2017). However, as proved by McDermott et al. (2016), the role of perception in the formulation of aesthetic judgment remains unclear. Therefore, musical consonance and dissonance remains a hotly debated topic (see Bowling et al., 2017), in need of further research to merge different approaches into a consistent theory.

## AUTHOR CONTRIBUTIONS

LT originally conceived the idea of the paper, elaborated the stimuli, provided all the figures, and significantly contributed to the results and discussion. NDS prepared the manuscript, co-authored the introduction and the results with LT, contributed to the discussion, wrote the conclusion, and finally revised the entire draft. AG wrote the section "Recurrence Quantification Analysis," reviewed the entire manuscript, and suggested useful ideas for the discussion. All authors equally contributed to the revision of the manuscript before agreeing on the final version.

## FUNDING

This work was funded by the Institute of Philosophy of Scientific and Technological Practice, Campus Bio-Medico University of Rome, under a 2015 Grant on "Embodiment."

## ACKNOWLEDGMENTS

LT is grateful to Universitat Pompeu Fabra (UPF) of Barcelona for access to their library.

## REFERENCES

Abel, M., Ahnert, K., and Bergweiler, S. (2009). Synchronization of sound sources. *Phys. Rev. Lett.* 103:114301. doi: 10.1103/PhysRevLett.103.114301

Benade, A. H. (1973). The physics of brasses. *Sci. Am.* 229, 24–35. doi: 10.1038/scientificamerican0773-24

Bidelman, G. M., and Heinz, M. G. (2011). Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing. *J. Acoust. Soc. Am.* 130, 1488–1502. doi: 10.1121/1.3605559

Bidelman, G. M., and Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J. Neurosci.* 29, 13165–13171. doi: 10.1523/JNEUROSCI.3900-09.2009

Bowling, D. L., Hoeschele, M., Kamraan, Z. G., and Tecumseh Fitch, W. (2017). The nature and nurture of musical consonance. *Music Percept.* 35, 118–121. doi: 10.1525/mp.2017.35.1.118

Bowling, D. L., and Purves, D. (2015). A biological rationale for musical consonance. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11155–11160. doi: 10.1073/pnas.1505768112

Cartwright, J. H. E., Douthettb, J., González, D. L., Krantzd, R., and Piro, O. (2010). Two musical paths to the Farey series and devil's staircase. *J. Math. Music* 4, 57–74. doi: 10.1080/17459737.2010.485001

Cartwright, J. H. E., Gonzalez, D. L., and Piro, O. (2001). Pitch perception: a dynamical-systems perspective. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4855–4859. doi: 10.1073/pnas.081070998

Cartwright, J. H. E., Gonzalez, D. L., Piro, O., and Stanziali, D. (2002). Aesthetics, dynamics, and musical scales: a golden connection. *J. New Music Res.* 31, 51–58. doi: 10.1076/jnmr.31.1.51.8099

Cross, I. (2003). Music as a biocultural phenomenon. *Ann. N. Y. Acad. Sci.* 999, 106–111. doi: 10.1196/annals.1284.010

Di Stefano, N., Focaroli, V., Giuliani, A., Formica, D., Taffoni, F., and Keller, F. (2017). A new research method to test auditory preferences in young listeners: results from a consonance versus dissonance perception study. *Psychol. Music* 45, 699–712. doi: 10.1177/0305735616681205

Eckmann, J. P., Kamphorst, S. O., and Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhys. Lett.* 4, 973–976. doi: 10.1209/0295-5075/4/9/004

Fastl, H., and Zwicker, E. (2006). *Psychoacoustic. Facts and Models*. Berlin: Springer.

Foo, F., King-Stephens, D., Weber, P., Laxer, K., Parvizi, J., and Knight, R. T. (2016). Differential processing of consonance and dissonance within the human superior temporal gyrus. *Front. Hum. Neurosci.* 10:154. doi: 10.3389/fnhum.2016.00154

Frova, A. (1999). *Fisica nella Musica*. Bologna: Zanichelli.

González-García, N., González, M. A., and Rendón, P. L. (2016). Neural activity related to discrimination and vocal production of consonant and dissonant musical intervals. *Brain Res.* 1643, 59–69. doi: 10.1016/j.brainres.2016.04.065

Helmholtz, H. (1954). *On the Sensations of Tone*. New York, NY: Dover Publications.

Kameoka, A., and Kuriyagawa, M. (1969a). Consonance theory part I: consonance of dyads. *J. Acoust. Soc. Am.* 45, 1451–1459. doi: 10.1121/1.1911623

Kameoka, A., and Kuriyagawa, M. (1969b). Consonance theory part II: consonance of complex tones and its calculation method. *J. Acoust. Soc. Am.* 45, 1460–1469. doi: 10.1121/1.1911624

Kennel, M. B., Brown, R., and Abarbanel, H. D. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* 45, 3403. doi: 10.1103/PhysRevA.45.3403

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., and Schlaug, G. (2005). Adults and children processing music: An fMRI study. *Neuroimage* 25, 1068–1076. doi: 10.1016/j.neuroimage.2004.12.050

Koelsch, S., and Mulder, J. (2002). Electric brain responses to inappropriate harmonies during listening to expressive music. *Clin. Neurophysiol.* 113, 862–869. doi: 10.1016/S1388-2457(02)00050-0

Large, E. W., and Almonte, F. V. (2012). Neurodynamics, tonality, and the auditory brainstem response. *Ann. N. Y. Acad. Sci.* 1252, E1–E7. doi: 10.1111/j.1749-6632.2012.06594.x

Large, E. W., and Tretakis, A. E. (2005). Tonality and nonlinear resonance. *Ann. N. Y. Acad. Sci.* 1060, 53–56. doi: 10.1196/annals.1360.046

Lohri, A. (2016). *Kombinationstöne und Tartinis "terzo suono"*. Mainz: Schott Music.

Lots, I. S., and Stone, L. (2008). Perception of musical consonance and dissonance: an outcome of neural synchronization. *J. R. Soc. Interface* 5, 1429–1434. doi: 10.1098/rsif.2008.0143

Manetti, C., Ceruso, M. A., Giuliani, A., Webber, C. L. Jr., and Zbilut, J. P. (1999). Recurrence quantification analysis as a tool for characterization of molecular dynamics simulations. *Phys. Rev. E* 59, 992–998. doi: 10.1103/PhysRevE.59.992

Marwan, N., Romano, M. C., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Phys. Rep.* 438, 237–329. doi: 10.1016/j.physrep.2006.11.001

McCauley, J. L. (1994). *Chaos, Dynamics and Fractals*. Cambridge: Cambridge University Press.

McDermott, J., Schultz, A. F., Undurraga, E. A., and Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variations in music perception. *Nature* 535, 547–550. doi: 10.1038/nature18635

Minati, L., Rosazza, C., D'Incerti, L., Pietrocini, E., Valentini, L., Scaioli, V., et al. (2008). FMRI/ERP of musical syntax: comparison of melodies and

unstructured note sequences. *Neuroreport* 19, 1381–1385. doi: 10.1097/WNR.0b013e32830c694b

Nikolsky, A. (2015). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.01405

Orsucci, F., Giuliani, A., Webber, C., Zbilut, J., Fonagy, P., and Mazza, M. (2006). Combinatorics and synchronization in natural semiotics. *Phys. A Stat. Mech. Appl.* 361, 665–676. doi: 10.1016/j.physa.2005.06.044

Pankovski, T., and Pankovska, E. (2017). Emergence of the consonance pattern within synaptic weights of a neural network featuring Hebbian neuroplasticity. *Biol. Insp. Cong. Arch.* 22, 82–94. doi: 10.1016/j.bica.2017.09.001

Park, J. Y., Park, H., Kim, J., and Park, H. J. (2011). Consonant chords stimulate higher EEG gamma activity than dissonant chords. *Neurosci. Lett.* 488, 101–105. doi: 10.1016/j.neulet.2010.11.011

Parncutt, R., and Hair, G. (2011). Consonance and dissonance in theory and psychology: disentangling dissonant dichotomies. *J. Interdiscip. Music Stud.* 5, 119–166.

Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., et al. (2010). Functional specializations for music processing in the human newborn brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4758–4763. doi: 10.1073/pnas.0909074107

Piana, G. (2007). *Barlumi per una Filosofia Della Musica*. Morrisville: Lulu Press.

Plomp, R. (1976). *Aspects of Tone Sensation: A Psychophysical Study*. London: Academic Press.

Plomp, R., and Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* 38, 548–560. doi: 10.1121/1.1909741

Roads, C. (2001). *Microsound*. Cambridge, MA: The MIT Press.

Roederer, J. G. (2008). *The Physics and Psychophysics of Music*. New York, NY: Springer.

Schroeder, M. (1990). *Fractals, Chaos, Power Laws*. New York, NY: W.H Freeman and Co.

Schwartz, D. A., Howe, C. Q., and Purves, D. (2003). The statistical structure of human speech sounds predicts music universals. *J. Neurosci.* 23, 7160–7168.

Serra, J., Serra, X., and Andrzejak, R. G. (2009). Cross recurrence quantification for cover song identification. *New J. Phys.* 11:093017. doi: 10.1088/1367-2630/11/9/093017

Trulla, L. L., Giuliani, A., Zbilut, J. P., and Webber, C. L. (1996). Recurrence quantification analysis of the logistic equation with transients. *Phys. Lett. A* 223, 255–260. doi: 10.1016/S0375-9601(96)00741-4

Trulla, L. L., Giuliani, A., Zimatore, G., Colosimo, A., and Zbilut, J. P. (2005). Non-linear assessment of musical consonance. *Electron. J. Theor. Phys.* 8, 22–34.

Wang, X. (2013). The harmonic organization of auditory cortex. *Front. Syst. Neurosci.* 7:114. doi: 10.3389/fnsys.2013.00114

Webber, C. L., and Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* 76, 965–973. doi: 10.1152/jappl.1994.76.2.965

Zimatore, G., Giuliani, A., Hatzopoulos, S., Martini, A., and Colosimo, A. (2003). Otoacoustic emissions at different click intensities: invariant and subject-dependent features. *J. Appl. Physiol.* 95, 2299–2305. doi: 10.1152/japplphysiol.00667.2003

Zimatore, G., Hatzopoulos, S., Giuliani, A., Martini, A., and Colosimo, A. (2002). Comparison of transient otoacoustic emission (TEOAE) responses from neonatal and adult ears. *J. Appl. Physiol.* 92, 2521–2528. doi: 10.1152/japplphysiol.01163.2001

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2018 Trulla, Di Stefano and Giuliani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Neuroscience

CONCEPTUAL ANALYSIS
published: 16 April 2018
doi: 10.3389/fnins.2018.00246

# Music Evolution in the Laboratory: Cultural Transmission Meets Neurophysiology

*Massimo Lumaca*[1]*, *Andrea Ravignani*[2,3,4] *and Giosuè Baggio*[5]

[1] Center for Music in the Brain, Department of Clinical Medicine, Aarhus University and The Royal Academy of Music Aarhus/Aalborg, Aarhus, Denmark, [2] Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium, [3] Research Department, Sealcentre Pieterburen, Pieterburen, Netherlands, [4] Language and Cognition Department, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, [5] Language Acquisition and Language Processing Lab, Department of Language and Literature, Norwegian University of Science and Technology, Trondheim, Norway

In recent years, there has been renewed interest in the biological and cultural evolution of music, and specifically in the role played by perceptual and cognitive factors in shaping core features of musical systems, such as melody, harmony, and rhythm. One proposal originates in the language sciences. It holds that aspects of musical systems evolve by adapting gradually, in the course of successive generations, to the structural and functional characteristics of the sensory and memory systems of learners and "users" of music. This hypothesis has found initial support in laboratory experiments on music transmission. In this article, we first review some of the most important theoretical and empirical contributions to the field of music evolution. Next, we identify a major current limitation of these studies, i.e., the lack of direct *neural* support for the hypothesis of cognitive adaptation. Finally, we discuss a recent experiment in which this issue was addressed by using event-related potentials (ERPs). We suggest that the introduction of neurophysiology in cultural transmission research may provide novel insights on the micro-evolutionary origins of forms of variation observed in cultural systems.

**Keywords: cultural transmission, diffusion chains, signaling games, iterated learning, music universals, music diversity, neural predictors, mismatch negativity (MMN)**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky, Independent Researcher, Los Angeles, CA, United States*

***Reviewed by:***
*McNeel Gordon Jantzen, Western Washington University, United States*
*Laura Verga, Maastricht University, Netherlands*
*Vera Kempe, Abertay University, United Kingdom*
*Seana Coulson, University of California, San Diego, United States*

***Correspondence:**
*Massimo Lumaca*
*massimo.lumaca@gmail.com*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *21 September 2017*
***Accepted:*** *29 March 2018*
***Published:*** *16 April 2018*

***Citation:***
*Lumaca M, Ravignani A and Baggio G (2018) Music Evolution in the Laboratory: Cultural Transmission Meets Neurophysiology. Front. Neurosci. 12:246. doi: 10.3389/fnins.2018.00246*

## INTRODUCTION

There has recently been a surge of interest in the biological and cultural origins, and evolution of music (Wallin et al., 2001; McDermott and Hauser, 2005; Patel, 2010). Music is prominent in virtually all human societies, and in its most sophisticated versions it is only attested in humans. This fact raises two important questions: how did music originate? And how did it evolve in its current forms? One intriguing issue here, especially in relation to the cognitive and neural bases of music evolution (Honing et al., 2015), is that of the *evolution of musical structure*. Musical systems are structured at several levels, from melody and harmony to rhythm and composition, in ways that may resemble the organization of other human generative systems, such as language (Jackendoff and Lerdahl, 1983; Jackendoff, 2009). The analogy between language and music may be pushed further, if one considers aspects of music that may be understood "semantically." Listening to music can evoke a wide range of extra-musical experiences, from emotional feelings (e.g., the sadness suggested by Albinoni's Adagio in G minor) to the mental imagery of specific referents

(e.g., characters or ideas in Wagnerian *Leitmotifs*) (Patel, 2010). Musical structures can and often do relate to a world of possible experiences and non-musical phenomena (Lerdahl, 2003) *expressively* (by being associated to internal affective states, e.g., emotional qualities), if not *representationally* (via relations of reference and truth, as language does) (Patel, 2010).

In this work, we focus on the cultural origins of musical syntax: the set of principles governing the combination of melody and rhythm into "well-formed" sequences (for a discussion on the evolution of semantic structures see Lumaca and Baggio, 2017, 2018; Ravignani and Verhoef, 2018). Some aspects of musical syntax, such as the organization of temporal structure and pitch intervals, display widespread distribution and striking cross-cultural similarities. For example, the tendency to use small intervals in non-polyphonic melodic phrases, or "proximity," has been observed across several musical traditions of the world, including indigenous tunes from North America, Europe, and Asia (Dowling, 1968; Von Hippel, 2000). Despite some exceptions, such as Scandinavian and Swiss yodeling music, proximity is a prominent feature of melodic structure. These shared attributes are known as "musical universals." Nevertheless, their form and frequency differ across and within different musical traditions of the world (Lomax, 1977; Rzeszutek et al., 2012; Savage et al., 2015). How can we explain both the invariance and the variation of structure in music? Which processes underlie the cross-cultural convergence toward common music traits or their diversification? In this paper, we suggest that neuroscience can provide critical methodological and theoretical tools for testing and generating hypotheses on this complex matter.

This article is organized as follows. We start by presenting a recent theoretical perspective in which music is understood as an evolving cultural system, adapting to the human brain [sections Linking Biological and Cultural Levels of Analysis and From Cultural Transmission to Neurophysiology (and Back)]. In section The Cognitive Level: Diffusion Chains and the Evolution of Musical Regularities in the Lab, we describe studies that support this view using data from behavioral experiments. In section The Neural Level: Constraints Imposed by a Neuronal Niche Drive the Emergence of Regularities, we transpose our analysis of cultural adaptation to the neural level. Partly using the "neuronal recycling hypothesis" as a theoretical framework (Dehaene and Cohen, 2007), we argue that music can adapt to a "neuronal niche" defined by the specific information processing constraints imposed by neural circuits originally evolved for auditory streaming.

To our knowledge, no one until recently has investigated this hypothesis by means of brain imaging or neurophysiology. In section Neural Predictors in Cultural Evolution Research, we describe a recent experiment in which this hypothesis was tested combining behavioral and neurophysiological methods. Finally (section The Neural Origins of Cultural Variation), we suggest that the introduction of concepts and methods from neuroscience in music evolution, and cultural evolution in general, can provide new insights on the process of cultural variation.

## Linking Biological and Cultural Levels of Analysis

Music may be seen as a complex adaptive system, shaped by various biological, environmental, and cultural factors. This has made it difficult for musicologists and cognitive scientists to analyze the evolutionary origins of musical structure. The predominant view during the last century was the *cultural account*, where music was seen as an entirely socio-cultural construct, free to vary with virtually no biological and environmental constraints on its structure and content (Nettl, 1983; Repp, 1991; Blacking et al., 1995). The striking diversity of musical forms, as attested across and within cultures, and over human history, seems to support this notion (Lomax, 1968; Henry, 1976). Yet, this account has been challenged by experiments in psychology and neuroscience, together supporting a broadly *biological account* of the origins of music. Several studies point to the existence of perceptuo-cognitive biases and constraints in music processing and production (e.g., Trehub, 2000; Drake and Bertrand, 2001; Zatorre, 2001; Peretz and Zatorre, 2005; Deutsch, 2012) with some parallels in other species (Fitch, 2015). On this view, prototypical properties of music, such as a relatively steady beat, smooth melodic contours, tonality, and a narrow distance between adjacent tones (or "pitch proximity"), derive from built-in functional properties of the brain (McDermott and Hauser, 2005), which tend to manifest themselves in most human cultures (Lerdahl, 1992; Savage et al., 2015).

A recent view is that neither the "cultural account" nor the "biological account" can independently provide a satisfactory theory of the origins and evolution of musical structure (Trainor, 2015). Cultural accounts typically focus on the evolution of *musical systems*, while biological accounts investigate the evolution of the *human capacity to perceive, appreciate, and produce music* (also including musicality; Honing et al., 2015). These different accounts, however, may be connected within a more complete explanatory framework, if one accepts that music is neither an entirely arbitrary cultural construct nor strictly a biological product. Much like natural language, music is a cultural construct, which nonetheless rests upon, and is partly shaped by, human neurobiology. Our neurobiological makeup determines the scope and constraints of human auditory memory capacity, hierarchical sequence processing, attention, perceptual hearing threshold, and auditory scene analysis (Snyder, 2008; Deutsch, 2012). This is now a central tenet in the field of music cognition, and it is becoming increasingly accepted in cultural analyses of music, too. The open question is *how* neurobiological capacities, biases and constraints manifest themselves in actual musical systems (Trainor, 2015).

## From Cultural Transmission to Neurophysiology (and Back)

Answering this question requires theories, models, and empirical data that can effectively bridge the gap between the classical chasms of (cultural) evolutionary science: between individual-level and population-level processes, micro-evolutionary and macro-evolutionary processes (Mesoudi, 2011). Specifically, one

important question is how the individual's neurobiological endowment manifests itself in music at the population level. This issue was already known in linguistics as the "problem of linkage" (Kirby, 1999). A possible answer is "through cultural transmission." Music, much like language, is not only a richly structured symbolic system, but also a set of behaviors that is maintained over time by intergenerational transmission (Morley, 2013; Le Bomin et al., 2016).

During intergenerational transmission, cultural information must survive a "memory bottleneck" (Deacon, 1997): the set of all neurobiological biases or constraints that bind our capacity to infer (and store) the "rules" that govern a system of information[1]. The properties of the cultural system that fit best the human neurobiological filter—e.g., those that make information easier to process, encode, and recall—will have greater likelihood of being passed on to the next generation. If this view is correct, in the long run the neurobiological endowment of individuals should be reflected in the musical corpus at the population-level.

This view of transmission, emphasizing *adaptation of fast-changing cultural systems to a largely stable neurocognitive architecture*, was developed in evolutionary linguistics to account for the emergence of structure in human languages, including putative linguistic universals (Christiansen and Chater, 2008). Recent methodological advances (Mesoudi, 2015; Edmiston et al., 2018) have provided support for this view in controlled laboratory conditions. In most experiments, groups of individuals engage in simple, controlled forms of knowledge transmission, for example from a participant (a sender) to another (a receiver), along a diffusion chain. Each participant represents a "generation," and each interaction between participants allows for the passage of information across generations (Esper, 1925; Bartlett, 1932). The set of items transmitted along a diffusion chain (e.g., linguistic or musical phrases) is a finite sample drawn from the (infinite) set of items that learners have to generalize from. A challenge for research on cultural transmission is to show that core properties of the artificial systems being transmitted are also properties of the actual cultural systems being modeled and that the mechanisms at work in artificial conditions are also at work in real cultural evolution. In a landmark study, Kirby et al. (2008) showed how miniature "languages" emerge in the course of transmission from initial random associations of signals and meanings. When these pairings are transmitted across "generations" of participants, some regularities emerge, including *compositionality* (Hockett, 1960), as observed in human language. This result supports the view that core properties of language can be explained by the interplay of individual cognitive biases (sensu Brighton et al., 2005) and iterated cultural learning and transmission. Recent studies on animal models of cultural learning further support this conclusion (e.g., for non-human primates see Claidière et al., 2014; for a seminal study on zebra finches see Fehér et al., 2009).

One way to start bridging this gap in the musical domain, is to assume that music, like language, is a complex adaptive cultural system, shaped for thousands of years by cycles of transmission, acquisition, and use (Morley, 2013). Following this view, neurobiological biases and constraints, as discussed above, brought out through cultural transmission, would exert effects on the form and structure of music (Merker et al., 2015; Trainor, 2015; Mehr et al., 2018). This mechanism could explain some properties of temporal (rhythm, meter) and spectral (melody, harmony) dimensions of musical structure, which are likely to be the result of *adaptations* to the combined pressures of neural constraints and various socio-cultural forces (Merker, 2006; Merker et al., 2015; Trainor, 2015). This would in principle apply to both *invariants*—putative cultural *universals* shared by musical systems or traditions (Savage et al., 2015)—and *variation* among individuals, generations, and traditions.

This point is not new. Lévi-Strauss (1960) had already observed that some structural regularities observed across cultures (e.g., the fact that symbolic material tends to be organized in binary oppositions) are reflections of principles of brain organization. Therefore, neuroscience is expected to contribute to explanations of the emergence and evolution of structural regularities, including their convergence and diversity. However, to date this issue has been addressed only by behavioral studies, and only to explain some invariant aspects of musical structure. In the next section, we summarize three of these lines of experimental work in the field of music evolution.

## THE COGNITIVE LEVEL: DIFFUSION CHAINS AND THE EVOLUTION OF MUSICAL REGULARITIES IN THE LAB

In recent experiments, a diffusion chain method was used to study how music evolves in the lab (Ravignani et al., 2016). This study aimed to test whether human psychological biases, amplified by cultural transmission, can explain the emergence of rhythmic universals (Trehub, 2015). In this experiment, participants were given a drumstick and an electronic drum pad. Participants in the first generation listened to 32 randomly generated, hence a-rhythmic, patterns of beats (the input), and were asked to reproduce each of them to the best of their abilities (the output). The "imperfect" output produced by this first generation of participants became the input for the next generation, whose task was to perform the rhythm they heard, and so on, along a diffusion chain. This paradigm is known as "iterated learning" (IL) (Kirby et al., 2008). Given the difficulty to memorize these patterns, errors were introduced in the emerging system of drumming sequences, slightly modifying the original patterns at each generation. Across generations, patterns became increasingly structured and easier to learn. After 8 generations, at the end of each diffusion chain, patterns showed regularities similar to those found across musical traditions of the world. These universal rhythmic regularities included a tendency toward small integer ratios (e.g., 1:1 and 2:1) of intervals between beat

[1]Our definition of "memory bottleneck" includes constraints on perceptual grouping; capacity and temporal limits of auditory memory, serial processing, and attention; constraints on the neurodynamics of the auditory system; perceptual hearing thresholds. We limited this list to constraints "directly" related to basic aspects of perception and cognition. We acknowledge that constraints of a different nature might have a formative power over musical structures (e.g., motoric, motoric-expressive, physiological, cross-modal, and semantics).

durations, and a relatively steady beat, also termed "isochrony" (Savage et al., 2015). This study represents the very first attempt to "grow" musical universals in the lab (Fitch, 2017), and sheds light on the cognitive and cultural mechanisms underlying the creation and vertical transmission of music (Le Bomin et al., 2016).

An IL study by Verhoef (2012) investigated the cultural evolution of combinatorial structures in musical systems. Participants were first exposed to a set of 12 whistles that they had to imitate immediately after listening by using a slide whistle (training phase). Next, they were asked to reproduce the whole set of signals as they remembered it (recall phase). The sequences generated by a participant were used to train the next one in the diffusion chain, and so on, until the end of the chain. In the course of transmission, structural regularities emerged, as predicted by previous computer simulations (de Boer, 2000). In the last generations, fewer discrete units were reused by individuals in concatenations, repetitions, or mirror forms to produce the entire vocabulary of whistles. Combinatoriality is a "design feature" of human language (Hockett, 1960) and it applies to musical structure, too. For instance, the authors observed that two distinct whistles were often combined into a single pattern by the next generation of individuals. Also, participants tended to produce mirror forms out of single patterns, so that more elements were shared between signals of the same set. With fewer units to memorize, organized in this manner, the set of signals was more structured, more compressed, and easier to learn and reproduce.

A more recent attempt to study music evolution in the lab is the work by Lumaca and Baggio (2017). The authors used a different model of cultural transmission than IL: multi-generational signaling games (MGSGs) (Moreno and Baggio, 2015; Nowak and Baggio, 2016). MGSGs are in essence an iterated variant of signaling games (Lewis, 1969; Skyrms, 2010) that combine basic aspects of semiotic models of coordination and communication (e.g., horizontal transmission; Galantucci and Garrod, 2011) with the intergenerational transmission of IL (Kirby et al., 2008). Two-person signaling games were organized in diffusion chains of 8 generations each. In each game, the sender and receiver were expected to converge, through repeated interactions, on a common code: a signaling system where 5 isochronous melodic riffs were associated to basic or compound emotions. This design can contribute to model different aspects of music transmission: first, a degree of alignment of internal states between musical senders (e.g. composers) and receivers (e.g., an audience) at two main levels, the *structural* and *affective* (Temperley, 2004; Bharucha et al., 2011); second, a partial asymmetry in information flow from senders to receivers, which is present in language and music transmission (e.g., from composers to listeners, from teachers to pupils, etc.). In each signaling trial, the sender was presented on the screen with one of the 5 equiprobable emotions (visualized as human facial expressions) and was asked to compose a 5-note isochronous riff on the computer keyboard. The receiver, after he listened to the riff via headphones, was asked to choose one of the 5 expressive faces displayed on the screen (i.e., the one possibly seen by the sender). A feedback was then presented simultaneously to both participants' screens, showing the expressive face seen by the sender and the one chosen by the receiver for the same melodic signal. This procedure was repeated at each successive trial. At the end of the game, the receiver (generation $n$) became the sender in the next game, with the same structure and a new participant as a receiver (generation $n + 1$), and so on, until the chain was completed. Senders were always asked to transmit the code they had learned in the previous game. Therefore, recall errors in the melodic signals (possibly "innovations") were introduced throughout the experiment. The authors observed the gradual evolution over generations of several structural features of musical phrases: pitch proximity and continuity, symmetry, and motivic structure.

Despite differences in their assumptions and methods, those three experiments have reached similar conclusions: the immediate effects of psychological constraints on the musical systems may be *weak*, but they are *amplified* in the course of inter-generational transmission (Boyd and Richerson, 1988; Kalish et al., 2007; Kirby et al., 2007; Thompson et al., 2016) or iterated reproduction (Jacoby and McDermott, 2017), leading the evolution of musical structures along non-random paths. If principles of auditory organization and memory constraints operate in similar ways also in the production and perception of actual music, they could similarly shape the evolution of historical systems in the course of iterated transmission. Convergence toward some of the musical structures found across populations (Savage et al., 2015) could be then explained, to some extent, by adaptation to a special niche, constituted by a restricted set of low-level perceptual and memory processes. In the rest of the paper we will refer to this special niche as "neuronal niche" (Dehaene and Cohen, 2007).

## THE NEURAL LEVEL: CONSTRAINTS IMPOSED BY A NEURONAL NICHE DRIVE THE EMERGENCE OF REGULARITIES

In recent years, there has been an increasing interest in how the brain accommodates and shapes novel cultural symbolic systems (Dehaene and Cohen, 2007). A leading hypothesis is that some cortical circuits, initially evolved as a result of specific selective pressures, are later "recycled" to accommodate novel cultural functions (Dehaene and Cohen, 2007; Simon et al., 2013; Dehaene et al., 2015; Skeide et al., 2017). Therefore, the acquisition of novel functions is constrained, however weakly, by prior human evolution. Once "culturally recycled," pre-existing systems and mechanisms maintain some of their original capacities and limitations, providing a neuronal niche within which culture may adapt and evolve. This also means that the variability observed in cultural systems is limited by brain structure and function across individuals and groups.

If this hypothesis is correct, near-universal characteristics of music (Savage et al., 2015) may be traced back to the computational infrastructure of human auditory cortex and other (e.g., motor, attentional etc.) areas of the brain. Trainor (2015) related the origins of certain invariant musical features as adaptations to bottom-up neural mechanisms of auditory

scene analysis (ASA), such as the sequential sound segregation and integration of within-stream elements (Bregman, 1994). These specific mechanisms have evolved specifically to detect and localize multiple sources of auditory objects and to extract regularities from the acoustic environment. They often involve the perceptual grouping of single-event auditory stimuli into auditory streams and operate following Gestalt principles of proximity, similarity, and continuity (Deutsch, 1999). They are automatic (pre-attentive), they emerge early in human development (Demany, 1982; Winkler et al., 2003), and they are widely conserved across species (Fay, 2008). This point shows that the ASA neural circuitry is likely phylogenetically older than human music. Thus, the exaptation (or evolutionary re-use) (Gould and Vrba, 1982) of this more ancient biological mechanism by music should impose constraints on the way music is stored and organized in the brain, and accordingly, on the way it is recalled during transmission. In this regard, perceptual and memory recall advantages have been reported for tone streams that conform to Gestalt principles of organization (Bendixen et al., 2010; Loui, 2012; Rohrmeier and Cross, 2013). The cross-cultural tendency to organize music following these principles (Huron, 2001), in addition to the findings reported by cultural transmission research (Verhoef, 2012; Ravignani et al., 2016; Lumaca and Baggio, 2017), may support the idea that the neurocomputational constraints of the human auditory system constitute a filter through which musical material must pass, adapt, and eventually evolve.

It is surprising that up until recently, no one has attempted to find (counter-) evidence of cultural adaptation using neural measures. Research has shown that even recently-encoded information is shaped by perceptual or memory constraints into more compressed and abstract forms (Tamariz, 2017). Yet, the neural mechanisms underlying this phenomenon remain unknown. One reason is arguably our limited understanding of how information is represented in the brain (Mesoudi et al., 2006). Current whole-brain methods, such as functional magnetic resonance (fMRI), are not well-suited to investigate the precise basis of mental representations (but see Haynes and Rees, 2006; Johnson and Johnson, 2014; Zadbood et al., 2017). Another issue is to establish a link between neural constraints on learning—neural activity underlying specific, fast, and accurate encoding processes (Sadtler et al., 2014)—and cultural adaptation. Electrophysiological methods, such as multi-unit recordings, seem ideal for this purpose, but they are too invasive to be performed on healthy individuals. Various animal models of social learning—in songbirds, primates, and other species—have provided useful information in this respect (Araki et al., 2016; Gadagkar et al., 2016; Tchernichovski and Lipkind, 2016). None of these species possesses cultural behaviors as rich and complex as human music. However, some of their behaviors exhibit structured patterns, which are maintained over time through inter-generational transmission. Cultural transmission, in turn, can shape animal vocal behavior so as to fit species-specific learning constraints (Fehér et al., 2009; Fitch, 2009).

The application of techniques and models used in language evolution allow researchers of animal behavior to explore the biology of culturally transmitted systems in simpler and more controlled conditions, and to answer questions about cultural adaptation that cannot be *directly* answered in humans using current methods (but see next section for *indirect* answers). For example, Araki et al. (2016) used cellular recordings to demonstrate the existence in zebra finches of constraints on neuronal temporal coding that limit song acquisition to certain species-specific temporal features. Juvenile birds acquire their songs by imitating adult tutors. Although zebra finches are not bound to learn only specific sequences, they do show significant consistencies in their vocal repertoires (Lachlan et al., 2016). Do these consistencies result from adaptation of song material to the zebra finch neural constraints on learning? Araki et al. (2016) found that a subset of neurons in the zebra finch auditory cortex responds synchronously and selectively to patterns of inter-syllable silent gap durations, which are typical of their songs. The same cell population was unresponsive to other species' songs. Temporal coding mechanism like this are thought to preserve the species-specific song identity from any random drifts that may be introduced during cultural transmission.

Critically, the same mechanisms might underlie learning behaviors that resemble cultural adaptation in humans. When presented with the songs of other species, zebra finches tend to gradually adjust the duration of inter-syllable intervals toward their own (species-specific) songs' temporal structures, in a way similar to the human adjustment of random auditory stimuli toward Gestalt features. To our knowledge, this work provides the first cellular-level support of the idea of a neurobiological basis of cultural adaptation. It remains to be determined to what extent their findings can be generalized to other species. Would similar neuronal constraints operate in humans? Could they explain perceptual predispositions for some musical features (e.g., for small intervals and isochronous beat)? Are those neuronal constraints species-specific or, instead, are they shared with other species (Nicolai et al., 2014)? Another critical question is whether inter-individual variability in the neural filter is reflected in forms of cultural variation, for example in participant behavior during transmission, or in the shape taken by cultural systems as a result of it. Cross-individual variability is typically regarded as a source of noise in cultural transmission research, and is often removed by means of various procedures. The idea of linking individual neural variability with cultural variation may lend itself well to investigations using brain imaging and electrophysiology, but no one until recently has adopted this approach in cultural transmission research.

## Neural Predictors in Cultural Evolution Research

In a recent experiment, Lumaca and Baggio (2016) addressed some of these issues using a neural predictors approach (Berkman and Falk, 2013). This entails use of neuroimaging (fMRI, PET) or electrophysiological methods (EEG/ERPs, MEG) to identify neural predictors of behavior (for examples in the music domain, see Golestani et al., 2002; Zatorre et al., 2012; Zatorre, 2013). Lumaca and Baggio (2016) used neural predictors of signaling behavior as a first approach to examine whether and how symbolic systems adapt to human neural information

processing systems, and to assess the effects of inter-individual variation in neural information processing on three core cultural behaviors: social learning, transmission, and regularization of signal sequences. To this purpose, the authors used one of the best-investigated brain signatures of auditory processing, the mismatch negativity (MMN) (Näätänen et al., 1978).

The MMN is a fronto-central negative wave, evoked by violations of some perceptual regularity (Paavilainen, 2013) which is picked up by the brain in a visual or auditory stimulus stream. The limited influence of attentive processes on the MNN (Paavilainen, 2013) and its onset (∼200 ms from the relevant stimulus) suggest that the MMN is a *low-level* marker of auditory processing. The encoding of regularities from an auditory input, possibly through the same ASA mechanisms reported above, is an antecedent condition for the elicitation of the MMN (Näätänen et al., 2001). The efficiency of these mechanisms is revealed by the MMN latency and amplitude (Näätänen et al., 1993; Tervaniemi et al., 2001). Larger amplitudes or shorter latencies are typically associated to more accurate representations of the input material and, thus, they are taken as proxies of more efficient encoding mechanisms. The MMN has been used to study how efficiently an individual's auditory system extracts and encodes regularities from acoustic inputs, and how this process may affect linguistic and musical behaviors. For example, differences in ERP responses in infants have been successfully used in various studies to predict cognitive and linguistic development (Molfese and Molfese, 1997; Choudhury and Benasich, 2011). Overall, these studies open up the possibility of using low-level neural markers to predict individual behavior during transmission and acquisition of language, music, and cultural material more generally. Structural properties of symbolic systems may thus be understood as adaptations to information processing bottlenecks during cultural transmission (Kirby, 2001; Tamariz and Kirby, 2015). It should then be possible, for example, to find a relationship between individual brain processing capabilities or limitations, and the degree of regularization imposed by each individual on the cultural material that is being transmitted and acquired.

Neurophysiological (ERP) evidence for this type of effect was provided by Lumaca and Baggio (2016) in the domain of melodic structure. The authors combined ERPs with diffusion chains on two successive days. On day 1, they identified a neural correlate of extracting regularities from 5-tone sequences in musically naïve individuals in a classical auditory oddball paradigm. ERPs were recorded while participants were presented with randomly interleaved standard (80%) and deviant (20%) stimuli: there was no task for the participants, who were watching a silent movie throughout the session. On day 2, participants played a reduced version of MGSGs, with melodic systems of the same kind used by Lumaca and Baggio (2017). Each participant played the first signaling game as receiver (learner) and the second as sender (transmitter)[2]. The main question addressed by the authors was whether constraints and biases on auditory processing could drive the melodic material toward known Gestalt principles of perceptual organization (Lumaca and Baggio, 2017). The results showed that inter-individual variation in neural information processing, as revealed by the latency of the MMN on day 1, predicted learning and transmission of melodic signaling systems in the MGSGs on day 2. Specifically, individuals with longer MMN latencies performed "worse" in the MGSGs, showing lower coordination, transmission, and accuracy. Yet, these participants introduced more innovations than participants with shorter MMN latencies. Inter-individual variation in neural auditory processing (or regularity encoding) may be sufficient to discriminate "better" from "worse" transmitters, as observed in the cultural transmission of music (Sawa, 2002). However, perhaps the most interesting finding was that participants with longer MMN latencies introduced more regularities in the artificial tone system, reproducing more often melodic structures that were more compressed (signals from the same set became more similar), more proximal (temporally adjacent elements in the signals were closer in pitch), and smoother (the sequences showed a coherent melodic direction) than the sequences they originally received. To our knowledge, this study is the first demonstration that three essential processes underlying cultural evolution (i.e., social learning, transmission, and innovation), and three near-universal properties of melodic structure (i.e., proximity, continuity, and compression) are constrained by the organization of sensory and memory systems in the brain. The MMN is only "the tip of the iceberg" here. The MMN is likely to reflect auditory scene analysis and encoding mechanisms. Constraints on these mechanisms, as revealed (among others) by MMN latencies, may represent a "neuronal niche" through which cultural material must pass, adapt, and evolve (see below). In a cultural evolutionary context, this finding may provide clues to the origins of forms of variation observed in cultural symbolic systems. We discuss this point in the next paragraph.

## The Neural Origins of Cultural Variation

Human cultural traits show a myriad different forms across world cultures. Music, like language, provides an excellent example of this diversity, within and between populations (Lomax, 1959; Rzeszutek et al., 2012). For instance, the tendency toward the use of intervals of small size or the division of the octave (2:1) into a limited number of tones (or "discreteness") as observed in several cultures (Merriam et al., 1956; Dowling, 1968) is counterbalanced by significant diversity, within and between those cultures, in the relative frequency of such traits (Savage et al., 2015). The frequency distribution of proximal intervals (<700 cents; Savage et al., 2015) differs across musical traditions, with variation being mostly confined to the interval range 0 (unison) to 6 semitones (Huron, 2001). A similar diversity was found in the "tonal material" of musical cultures (i.e., the total set of discrete pitches within an octave), which spans from the 12 semitones of the Western musical scale to the 22–24 microtonal steps of North Indian and Arabic scales (Malm, 1967; Ayari and McAdams, 2003).

[2]In signaling games with fixed roles, including all MGSGs, the receiver tends to learn the code transmitted by the sender. In other words, there is asymmetry in the division of coordination labor between the sender and the receiver, with most coordination work (most code changes) falling to the latter (Nowak and Baggio, 2016).

The evolutionary mechanisms that affect the relative frequency of musical characters, such as random cultural drifts and biased selection, have been extensively studied in recent years (Mesoudi, 2015). For example, MacCallum et al. (2012) used a biologically-inspired evolutionary system to explore the effects of "aesthetic" selection on the frequency distribution of musical characters. A population of listeners was asked to rate the pleasantness of randomly generated tunes. The top-rated tunes recombined or mutated into novel variants that were in turn evaluated by a new generation of consumers. The authors reported an over-time increase of characters classically regarded as "musical," such as isochrony and chordal clarity. This work was the first of its kind to show that consumers' preferences can deeply shape the evolution of music in the near absence of learning and memory pressures. It is still controversial whether aesthetic preferences are just a social construct, changing over time, or if instead they are themselves stable information processing biases (for an in-depth discussion on this topic see Hodges, 2009; Huron, 2009). In a recent model, Reber et al. (2004) combined the two proposals. Specifically, the authors put forward the hypothesis that aesthetic preferences result from an interaction between knowledge-dependent stylistic rules and information processing fluency for certain stimulus properties (e.g., symmetry, clarity, and the amount of information content) (Nieminen et al., 2011). This may explain the evolution of music toward specific features, such as symmetry and chordal clarity (MacCallum et al., 2012; Verhoef, 2012; Lumaca and Baggio, 2017). A similar proposal was made by Haiman (2011) to explain the emergence of symmetric compounds in language. These arguments are still hypothetical, but we are now starting to understand the *effects of these biases* on the cultural evolution of music (Savage and Brown, 2007). Specifically, we know that these processes can enhance the diversity of musical behaviors and forms, but they can also produce local homogeneity[3]. While those mechanisms can explain how musical variants spread over time in a population, the sources of variability remain to a large extent elusive.

Up until now, only four main mechanisms of variation have been considered in music: *creative innovation* (e.g., via original musical composition), *borrowing* (through blending or syncretism), *translation* (from one tonal system to another; Alekseyev, 1986), and *random mutation* (errors in music copying or performance) (Savage and Brown, 2007). Lumaca and Baggio (2016) provided evidence for an additional mechanism: *individual neural variability*. One could argue that every individual in a population represents a distinct and unique "neuronal niche" (Dehaene and Cohen, 2007), through which cultural material is filtered and to which it may eventually adapt. Minor inter-individual differences in neural information processing can manifest themselves in differences in musical behavior. Moreover, they can be amplified and spread via different cultural evolutionary mechanisms. Small differences in learning or information processing can have large system-level effects, if they are amplified by cultural transmission.

One tenet of cultural transmission research is that cultural systems evolve toward certain prior distributions, known as "cognitive attractors" or "inductive biases" (Sperber, 1996; Griffiths et al., 2008). Strong versions of this account have been challenged by recent modeling work (Navarro et al., 2017). The convergence toward priors holds in the (implausible) scenario where all learners are endowed with the same identical prior. However, when learners instantiate (slightly) different constraints, the emerging cultural systems may reflect the more idiosyncratic biases of *some* individuals. In light of our findings, one could suggest that individuals with "tighter bottlenecks" exert a disproportionately large effect on the evolution of musical structures (see Ravignani et al., 2018 for some issues concerning this view). Similarly, differences between populations in brain function and anatomy may, at least in part, be reflected in differences in the structure of the symbolic systems in use. This account has recently found some support in language evolution research. Dediu and Ladd (2007) have shown that the population-level frequencies of two human genes involved in brain growth, Microcephalin and ASPM, are reliably associated with the presence or absence of linguistic tones in that population. The authors' proposal is that variants of these genes may determine small biases at the individual level in the processing and acquisition of linguistic tones, which may in turn give rise to distinct language variants. Those variants are hardly detectable in individual subjects, because tonal and non-tonal languages can be acquired by any individual, independently of genetic variants (Ladd et al., 2008). But when their effects are amplified by inter-generational transmission (Kirby et al., 2008), these variants may give rise to measurable, large-scale population differences.

Dediu and Ladd (2007) is the first study suggesting that variation, as observed in cultural traits and in their distribution, may originate in interindividual neurogenetic variability. Lumaca and Baggio (2016) provide converging neurophysiological evidence in support of this view (for the genetic bases of inter-individual variation in musicality, see Gingras et al., 2015). Genetic and neural variability are not the only source of cultural variation, but they are likely to play a prominent role in any future theory of the biological roots of culture. For example, Brown et al. (2014) have shown that musical and genetic diversity may correlate to some degree. After sampling a set of traditional songs from 9 indigenous populations in Taiwan, they measured the relative distance for 41 properties of song structure and performance-style. Music and genetic distance among the populations were significantly correlated. A similar relation was found in Eurasian populations (Pamjav et al., 2012). The study of genetic and neural variability may help address questions that were considered taboo in ethnomusicology since fairly recently: for example, whether a causal relationship exists between the distribution of some gene variants and aspects of musical systems and behaviors (Jordania, 2006, p. 101; Nikolsky, 2015). Such a theory requires the synergic and coordinated effort of genetics, neuroscience, and research on cultural evolution. The recent drive toward a "grand synthesis" of the latter discipline (Brewer et al., 2017) makes this possibility somewhat more likely.

[3]The re-use of Wagner's musical ideas by other composers during Nazi Germany and the emergence and maintenance of stylistic clusters in contemporary pop music are clear examples of biased selection.

## CONCLUSIONS

In this paper, we have argued that some of the most fundamental (and still unresolved) issues in music evolution can be addressed using the methods of cognitive neuroscience. This approach so far suggests a novel hypothesis on the mechanisms behind forms of cultural variation in musical systems. This line of work can also shed light on the "problem of linkage" (Kirby, 1999). Up until recently, this problem has been framed at only two levels of explanation. At the *behavioral level*, individual behaviors (e.g., code changes) that serve coordination and communication are linked to population-level patterns. At the *cognitive level*, sensory or memory constraints in individuals are identified in order to account for properties (e.g., structural features) of cultural systems. We suggest that a third level, the *neural level*, should be taken into consideration when developing accounts of the origins and evolution of structure in cultural systems, as is the case for accounts of the organization and function of information processing systems (Marr, 1982; Baggio et al., 2012, 2014, 2016). Thus, we can address questions in the cultural domain such as: which sources produce cultural diversity (computational level); through which mechanisms it may arise (e.g., inter-individual variation; algorithmic level); and which physical substrates, if any, those mechanisms exploit (i.e., the human brain; implementational level). We believe that explanations at all three levels are necessary to understand human cultural transmission. This requires (1) analyzing the structural and dynamic properties of the cultural systems (or codes) themselves, (2) determining how those are shaped by perceptual and cognitive biases and constraints, and (3) identifying the biological roots of such biases and constraints using neural and genetic data. This proposal generates several new questions, such as: to what extent do neural processes drive cultural evolution? How does inter-individual variation in brain function and structure affect variation in cultural behaviors? How does the distribution of neural traits in a population affect the structure of the symbolic system itself? How do these traits interact with aesthetic processing biases and the environment at large in the cultural evolution of music? How specific and accurate can neuroprediction get in the context of cultural evolution? Here, we hope to have shown that these questions are worth asking, and are largely amenable to scientific inquiry.

## AUTHOR CONTRIBUTIONS

ML wrote the article. AR and GB made additional contributions and edited the manuscript. All authors approved the manuscript for publication.

## FUNDING

AR was supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665501 with the research Foundation Flanders (FWO) (Pegasus$^2$ Marie Curie fellowship 12N5517N awarded to AR), and a visiting fellowship in Language Evolution from the Max Planck Society (awarded to AR).

## ACKNOWLEDGMENTS

We are grateful to Monica Tamariz, Bruno Gingras, and Aleksey Nikolsky for their helpful comments during the revision of the manuscript. Center for Music in the Brain is funded by the Danish National Research Foundation (DNRF117).

## REFERENCES

Alekseyev, E. (1986). *Early Folkloric Intonation. Pitch Aspect [Раннефольклорное Интонирование: Звуковысотный Аспект]*. Moscow: Sovetskii Kompozitor.

Araki, M., Bandi, M. M., and Yazaki-Sugiyama, Y. (2016). Mind the gap: Neural coding of species identity in birdsong prosody. *Science* 354, 1282–1287. doi: 10.1126/science.aah6799

Ayari, M., and McAdams, S. (2003). Aural analysis of Arabic improvised instrumental music (taqsim). *Music Percept.* 21, 159–216. doi: 10.1525/mp.2003.21.2.159

Baggio, G., Stenning, K., and van Lambalgen, M. (2016). "Semantics and Cognition," in *The Cambridge Handbook of Formal Semantics*, eds M. Aloni and P. Dekker (Cambridge: Cambridge University Press), 756–774.

Baggio, G., van Lambalgen, M., and Hagoort, P. (2012). "Language, Linguistics and Cognition," in *Handbook of the Philosophy of Linguistics*, eds R. Kempson, T. Fernando, and N. Asher (Amsterdam, NL: Elsevier), 325–355.

Baggio, G., van Lambalgen, M., and Hagoort, P. (2014). Logic as Marr's computational level: four case studies. *Top. Cogn. Sci.* 7, 287–298. doi: 10.1111/tops.12125

Bartlett, F. C. (1932). *Remembering: An Experimental and Social Study*. Cambridge, MA: Cambridge University Press.

Bendixen, A., Denham, S. L., Gyimesi, K., and Winkler, I. (2010). Regular patterns stabilize auditory streams. *J. Acoust. Soc. Am.* 128, 3658–3666. doi: 10.1121/1.3500695

Berkman, E. T., and Falk, E. B. (2013). Beyond brain mapping: using neural measures to predict real-world outcomes. *Curr. Dir. Psychol. Sci.* 22, 45–50. doi: 10.1177/0963721412469394

Bharucha, J. J., Paroo, K., and Curtis, M. (2011). "Alignment of brain states: response to commentaries," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Oxford: Oxford University Press), 195–198.

Blacking, J., Byron, R., and Nettl, B. (1995). *Music, Culture, and Experience: Selected Papers of John Blacking*. Chicago, IL: University of Chicago Press.

Boyd, R., and Richerson, P. J. (1988). *Culture and the Evolutionary Process*. Chicago, IL: University of Chicago Press.

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Brewer, J., Gelfand, M., Jackson, J. C., MacDonald, I. F., Peregrine, P. N., Richerson, P. J., et al. (2017). Grand challenges for the study of cultural evolution. *Nat. Ecol. Evol.* 1:70. doi: 10.1038/s41559-017-0070

Brighton, H., Smith, K., and Kirby, S. (2005). Language as an evolutionary system. *Phys. Life Rev.* 2, 177–226. doi: 10.1016/j.plrev.2005.06.001

Brown, S., Savage, P. E., Ko, A. M. S., Stoneking, M., Ko, Y. C., Loo, J. H., et al. (2014). Correlations in the population structure of music, genes and language. *Proc. R. Soc. B.* 281:20132072. doi: 10.1098/rspb.2013.2072

Choudhury, N., and Benasich, A. A. (2011). Maturation of auditory evoked potentials from 6 to 48 months: prediction to 3 and 4 year

language and cognitive abilities. *Clin. Neurophysiol.* 122, 320–338. doi: 10.1016/j.clinph.2010.05.035

Christiansen, M. H., and Chater, N. (2008). Language as shaped by the brain. *Behav. Brain Sci.* 31, 489–489. doi: 10.1017/S0140525X08004998

Claidière, N., Smith, K., Kirby, S., and Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. *Proc. Biol. Sci.* 281:20141541. doi: 10.1098/rspb.2014.1541

Deacon, T. (1997). *The Symbolic Species.* New York, NY: W.W.Norton.

de Boer, B. (2000). Self-organization in vowel systems. *J. Phon.* 28, 441–465. doi: 10.1006/jpho.2000.0125

Dediu, D., and Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10944–10949. doi: 10.1073/pnas.0610848104

Dehaene, S., and Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron* 56, 384–398. doi: 10.1016/j.neuron.2007.10.004

Dehaene, S., Cohen, L., Morais, J., and Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nat. Rev. Neurosci.* 16, 234–244. doi: 10.1038/nrn3924

Demany, L. (1982). Auditory stream segregation in infancy. *Infant Behav. Dev.* 5, 261–276. doi: 10.1016/S0163-6383(82)80036-2

Deutsch, D. (1999). "Grouping mechanisms in music," in *The Psychology of Music 2nd Edn*, ed D. Deutsch (San Diego, CA: Academy Press), 99–134.

Deutsch, D. (2012). *The Psychology of Music 3rd edn.* San Diego, CA: Academy Press.

Dowling, W. J. (1968). Rhythmic fission and perceptual organization. *J. Acoust. Soc. Am.* 44:369. doi: 10.1121/1.1970461

Drake, C., and Bertrand, D. (2001). The quest for universals in temporal processing in music. *Ann. N. Y. Acad. Sci.* 930, 17–27. doi: 10.1111/j.1749-6632.2001.tb05722.x

Edmiston, P., Perlman, M., and Lupyan, G. (2018). Repeated imitation makes human vocalizations more word-like. *Proc. R. Soc. B* 285:20172709. doi: 10.1098/rspb.2017.2709

Esper, E. A. (1925). A technique for the experiment investigation of associative interference in artificial linguistic material. *Lang. Monographs* 1, 1–47.

Fay, R. R. (2008). "Sound source perception and stream segregation in nonhuman vertebrate animals," in *Auditory Perception and Sound Sources*, eds W. A. Yost, A. N. Popper, and R. R. Fay (New York, NY: Springer), 307–323.

Fehér, O., Wang, H., Saar, S., Mitra, P. P., and Tchernichovski, O. (2009). *De novo* establishment of wild-type song culture in the zebra finch. *Nature* 459, 564–568. doi: 10.1038/nature07994

Fitch, W. T. (2009). Animal behaviour: birdsong normalized by culture. *Nature* 459, 519–520. doi: 10.1038/459519a

Fitch, W. T. (2015). Four principles of bio-musicology. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:20140091. doi: 10.1098/rstb.2014.0091

Fitch, W. T. (2017). Cultural evolution: lab-cultured musical universals. *Nat. Hum. Behav.* 1:0018. doi: 10.1038/s41562-016-0018

Gadagkar, V., Puzerey, P. A., Chen, R., Baird-Daniel, E., Farhang, A. R., and Goldberg, J. H. (2016). Dopamine neurons encode performance error in singing birds. *Science* 354, 1278–1282. doi: 10.1126/science.aah6837

Galantucci, B., and Garrod, S. (2011). Experimental semiotics: a review. *Front. Hum. Neurosci.* 5:11. doi: 10.3389/fnhum.2011.00011

Gingras, B., Honing, H., Peretz, I., Trainor, L. J., and Fisher, S. E. (2015). Defining the biological bases of individual differences in musicality. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:20140092. doi: 10.1098/rstb.2014.0092

Golestani, N., Paus, T., and Zatorre, R. J. (2002). Anatomical correlates of learning novel speech sounds. *Neuron* 35, 997–1010. doi: 10.1016/S0896-6273(02)00862-0

Gould, S. J., and Vrba, E. S. (1982). Exaptation—a missing term in the science of form. *Paleobiology* 8, 4–15. doi: 10.1017/S0094837300004310

Griffiths, T. L., Kalish, M. L., and Lewandowsky, S. (2008). Review. Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3503–3514. doi: 10.1098/rstb.2008.0146

Haiman, J. (2011). "Competing motivations," in *The Oxford handbook of Linguistic Typology*, ed J. J. Song (Oxford, UK: Oxford University Press), 148–165.

Haynes, J.-D., and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7, 523–534. doi: 10.1038/nrn1931

Henry, E. O. (1976). The variety of music in a north Indian village: reassessing cantometrics. *Ethnomusicology* 20:49. doi: 10.2307/850820

Hockett, C. F. (1960). The origin of speech. *Sci. Am.* 203, 89–96. doi: 10.1038/scientificamerican0960-88

Hodges, D. A. (2009). "The neuroaesthetics of music," in *The Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford, UK: Oxford University Press), 247–262.

Honing, H., ten Cate, C., Peretz, I., and Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:20140088. doi: 10.1098/rstb.2014.0088

Huron, D. (2001). Tone and voice: a derivation of the rules of voice-leading from perceptual principles. *Music Percept.* 19, 1–64. doi: 10.1525/mp.2001.19.1.1

Huron, D. (2009). Aesthetics, in *The Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford, UK: Oxford University Press), 151–159.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Percept.* 26, 195–204. doi: 10.1525/mp.2009.26.3.195

Jackendoff, R. S., and Lerdahl, F. (1983). *A Generative Theory of Tonal Music.* Cambridge, MA: MIT Press.

Jacoby, N., and McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Curr. Biol.* 27, 359–370. doi: 10.1016/j.cub.2016.12.031

Johnson, M. R., and Johnson, M. K. (2014). Decoding individual natural scene representations during perception and imagery. *Front. Hum. Neurosci.* 8:59. doi: 10.3389/fnhum.2014.00059

Jordania, J. (2006). *Who Asked the First Question? The Origins of Human Choral Singing, Intelligence, Language and Speech*. Tbilisi: Logos.

Kalish, M. L., Griffiths, T. L., and Lewandowsky, S. (2007). Iterated learning: intergenerational knowledge transmission reveals inductive biases. *Psychon. Bull. Rev.* 14, 288–294. doi: 10.3758/BF03194066

Kirby, S. (1999). *Function, Selection, and Innateness: The Emergence of Language Universals*. Oxford, UK: Oxford University Press.

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Trans. Evol. Comput.* 5, 102–110. doi: 10.1109/4235.918430

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10681–10686. doi: 10.1073/pnas.0707835105

Kirby, S., Dowman, M., and Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5241–5245. doi: 10.1073/pnas.0608222104

Lachlan, R. F., van Heijningen, C. A. A., Ter Haar, S. M., and Ten Cate, C. (2016). Zebra finch song phonology and syntactical structure across populations and continents—a computational comparison. *Front. Psychol.* 7:980. doi: 10.3389/fpsyg.2016.00980

Ladd, D. R., Dediu, D., and Kinsella, A. R. (2008). Languages and genes: reflections on biolinguistics and the nature-nurture question. *Biolinguistics* 2, 114–126.

Le Bomin, S., Lecointre, G., and Heyer, E. (2016). The evolution of musical diversity: the key role of vertical transmission. *PLoS ONE* 11:e0151570. doi: 10.1371/journal.pone.0151570

Lerdahl, F. (1992). Cognitive constraints on compositional systems. *Contemp. Music Rev.* 6, 97–121. doi: 10.1080/07494469200640161

Lerdahl, F. (2003). Two Ways in which music relates to the world. *Music Theor. Spectr.* 25, 367–373. doi: 10.1525/mts.2003.25.2.367

Lévi-Strauss, C. (1960). Éloge de l'anthropologie. Inaugural Lecture at Collège de France.

Lewis, D. (1969). *Convention: A Philosophical Study.* Cambridge, MA: Harvard University Press.

Lomax, A. (1959). Folk song style. *Am. Anthropol.* 61, 927–954. doi: 10.1525/aa.1959.61.6.02a00030

Lomax, A. (1968). *Folk song Style and Culture*. Washington, DC: American Association for the Advancement of Science.

Lomax, A. (1977). Universals in song. *World Music* 19, 117–129.

Loui, P. (2012). Learning and liking of melody and harmony: further studies in artificial grammar learning. *Top. Cogn. Sci.* 4, 554–567. doi: 10.1111/j.1756-8765.2012.01208.x

Lumaca, M., and Baggio, G. (2016). Brain potentials predict learning, transmission and modification of an artificial symbolic system. *Soc. Cogn. Affect. Neurosci.* 11, 1970–1979. doi: 10.1093/scan/nsw112

Lumaca, M., and Baggio, G. (2017). Cultural transmission and evolution of melodic structures in multi-generational signaling games. *Artif. Life* 23, 406–423. doi: 10.1162/ARTL_a_00238

Lumaca, M., and Baggio, G. (2018). Signaling games and the evolution of structure in language and music: a reply to Ravignani and Verhoef (2018). *Artif. Life.*

MacCallum, R. M., Mauch, M., Burt, A., and Leroi, A. M. (2012). Evolution of music by public choice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12081–12086. doi: 10.1073/pnas.1203182109

Malm, W. P. (1967). *Music Cultures of the Pacific, the Near East, and Asia.* Englewood Cliffs, NJ: Prentice Hall.

Marr, D. (1982). "Visual information processing: the structure and creation of visual representations," in *Recognition of Pattern and Form*, ed D. Albrecht (Berlin: Springer), 59–87.

McDermott, J., and Hauser, M. (2005). The origins of music: innateness, uniqueness, and evolution. *Music Percept.* 23, 29–59. doi: 10.1525/mp.2005.23.1.29

Mehr, S. A., Singh, M., York, H., Glowacki, L., and Krasnow, M. M. (2018). Form and function in human song. *Curr. Biol.* 28, 356–368. doi: 10.1016/j.cub.2017.12.042

Merker, B. (2006). The uneven interface between culture and biology in human music (commentary). *Music Percept.* 24, 95–98. doi: 10.1525/mp.2006.24.1.95

Merker, B., Morley, I., and Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:20140095. doi: 10.1098/rstb.2014.0095

Merriam, A. P., Whinery, S., and Fred, B. G. (1956). Songs of a rada community in trinidad. *Antropos* 51, 157–174.

Mesoudi, A. (2011). *Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences*. Chicago, IL: University of Chicago Press.

Mesoudi, A. (2015). Cultural evolution: a review of theory, findings and controversies. *Evol. Biol.* 43, 481–497. doi: 10.1007/s11692-015-9320-0

Mesoudi, A., Whiten, A., and Laland, K. N. (2006). Towards a unified science of cultural evolution. *Behav. Brain Sci.* 29, 329–347. doi: 10.1017/S0140525X06009083

Molfese, D. L., and Molfese, V. J. (1997). Discrimination of language skills at five years of age using event-related potentials recorded at birth. *Dev. Neuropsychol.* 13, 135–156. doi: 10.1080/87565649709540674

Moreno, M., and Baggio, G. (2015). Role asymmetry and code transmission in signaling games: an experimental and computational investigation. *Cogn. Sci.* 39, 918–943. doi: 10.1111/cogs.12191

Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology, and the Origins of Musicality*. Oxford, UK: Oxford University Press.

Näätänen, R., Gaillard, A. W., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329. doi: 10.1016/0001-6918(78)90006-9

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., and Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *Neuroreport* 4, 503–506. doi: 10.1097/00001756-199305000-00010

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., and Winkler, I. (2001). "Primitive intelligence" in the auditory cortex. *Trends Neurosci.* 24, 283–288. doi: 10.1016/S0166-2236(00)01790-2

Navarro, D. J. A., Perfors, A., Kary, A., Brown, S., and Donkin, C. (2017). When extremists win: on the behavior of iterated learning chains when priors are heterogeneous. *CogSci*. 847–852.

Nettl, B. (1983). *The Study of Ethnomusicology*: *Twenty-nine Issues and Concepts*. Chicago, IL: University of Illinois Press.

Nicolai, J., Gundacker, C., Teeselink, K., and Güttinger, H. R. (2014). Human melody singing by bullfinches (*Pyrrhula pyrrula*) gives hints about a cognitive note sequence processing. *Anim. Cogn.* 17, 143–155. doi: 10.1007/s10071-013-0647-6

Nieminen, S., Istók, E., Brattico, E., Tervaniemi, M., and Huotilainen, M. (2011). The development of aesthetic responses to music and their underlying neural and psychological mechanisms. *Cortex* 47, 1138–1146. doi: 10.1016/j.cortex.2011.05.008

Nikolsky, A. (2015). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: Prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.01405

Nowak, I., and Baggio, G. (2016). The emergence of word order and morphology in compositional languages via multigenerational signaling games. *J. Lang. Evol.* 1, 137–150. doi: 10.1093/jole/lzw007

Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: a review. *Int. J. Psychophysiol.* 88, 109–123. doi: 10.1016/j.ijpsycho.2013.03.015

Pamjav, H., Juhasz, Z., Zalan, A., Nemeth, E., and Damdin, B. (2012). A comparative phylogenetic study of genetics and folk music. *Mol. Genet. Genomics*. 287, 337–349. doi: 10.1007/s00438-012-0683-y

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford, UK: Oxford University Press.

Peretz, I., and Zatorre, R. J. (2005). Brain organization for music processing. *Annu. Rev. Psychol.* 56, 89–114. doi: 10.1146/annurev.psych.56.091103.070225

Ravignani, A., Delgado, T., and Kirby, S. (2016). Musical evolution in the lab exhibits rhythmic universals. *Nat. Hum. Behav.* 1:0007. doi: 10.1038/s41562-016-0007

Ravignani, A., Thompson, B., Grossi, T., Delgado, T., and Kirby, S. (2018). Evolving building blocks of rhyth m: how human cognition creates music via cultural transmission. *Ann. N.Y. Acad. Sci.* doi: 10.1111/nyas.13610

Ravignani, A., and Verhoef, T. (2018). Which melodic universals emerge from repeated signaling games? *Artif. Life*. 24.

Reber, R., Schwarz, N., and Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Pers. Soc. Psychol. Rev.* 8, 364–382. doi: 10.1207/s15327957pspr0804_3

Repp, B. H. (1991). "Some cognitive and perceptual aspects of speech and music," in *Music, Language, Speech and Brain*, eds J. Sundberg, L. Nord, and R. Carlson (Stockholm: MacMillan Press), 257–268.

Rohrmeier, M., and Cross, I. (2013). Artificial grammar learning of melody is constrained by melodic inconsistency: Narmour's principles affect melodic learning. *PLoS ONE* 8:e66174. doi: 10.1371/journal.pone.0066174

Rzeszutek, T., Savage, P. E., and Brown, S. (2012). The structure of cross-cultural musical diversity. *Proc. Biol. Sci.* 279, 1606–1612. doi: 10.1098/rspb.2011.1750

Sadtler, P. T., Quick, K. M., Golub, M. D., Chase, S. M., Ryu, S. I., Tyler-Kabara, E. C., et al. (2014). Neural constraints on learning. *Nature* 512, 423–426. doi: 10.1038/nature13665

Savage, P. E., and Brown, S. (2007). Toward a new comparative musicology. *Anal. Approach. World Music* 2, 148–197.

Savage, P. E., Brown, S., Sakai, E., and Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8987–8992. doi: 10.1073/pnas.1414495112

Sawa, G. D. (2002). Oral transmission in Arabic music, past and present. *Oral Trad.* 4, 254–265.

Simon, G., Lanoë, C., Poirel, N., Rossi, S., Lubin, A., Pineau, A., et al. (2013). Dynamics of the anatomical changes that occur in the brains of schoolchildren as they learn to read. *PLoS ONE* 8:e81789. doi: 10.1371/journal.pone.0081789

Skeide, M. A., Kumar, U., Mishra, R. K., Tripathi, V. N., Guleria, A., Singh, J. P., et al. (2017). Learning to read alters cortico-subcortical cross-talk in the visual system of illiterates. *Sci. Adv.* 3:e1602612. doi: 10.1126/sciadv.1602612

Skyrms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford, UK: Oxford University Press.

Snyder, B. (2008). "Memory for music," in *Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford, UK: Oxford University Press), 107–117.

Sperber, D. (1996). *Explaining Culture: A Naturalistic Approach*. Oxford: Oxford University Press.

Tamariz, M. (2017). Experimental studies on the cultural evolution of language. *Annu. Rev. Linguist.* 3, 389–407. doi: 10.1146/annurev-linguistics-011516-033807

Tamariz, M., and Kirby, S. (2015). Culture: copying, compression, and conventionality. *Cogn. Sci.* 39, 171–183. doi: 10.1111/cogs.12144

Tchernichovski, O., and Lipkind, D. (2016). Encoding vocal culture. *Science* 354, 1234–1235. doi: 10.1126/science.aal3205

Temperley, D. (2004). Communicative pressure and the evolution of musical styles. *Music Percept.* 21, 313–337. doi: 10.1525/mp.2004.21.3.313

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., and Näätänen, R. (2001). Superior formation of cortical memory traces for melodic patterns in musicians. *Learn. Mem.* 8, 295–300. doi: 10.1101/lm.39501

Thompson, B., Kirby, S., and Smith, K. (2016). Culture shapes the evolution of cognition. *Proc. Natl. Acad. Sci. U.S.A.* 113, 4530–4535. doi: 10.1073/pnas.1523631113

Trainor, L. J. (2015). The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:20140089. doi: 10.1098/rstb.2014.0089

Trehub, S. E. (2000). "Human processing predispositions and musical universals," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 427–448.

Trehub, S. E. (2015). Cross-cultural convergence of musical features. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8809–8810. doi: 10.1073/pnas.1510724112

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Lang. Cogn.* 4, 357–380. doi: 10.1515/langcog-2012-0019

Von Hippel, P. (2000). Redefining pitch proximity: tessitura and mobility as constraints on melodic universals. *Music Percept.* 17, 315–327. doi: 10.2307/40285820

Wallin, N. L., Merker, B., and Brown, S. (2001). *The Origins of Music*. Cambridge, MA: MIT Press.

Winkler, I., Kushnerenko, E., Horváth, J., Ceponiene, R., Fellman, V., Huotilainen, M., et al. (2003). Newborn infants can organize the auditory world. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11812–11815. doi: 10.1073/pnas.2031891100

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., and Hasson, U. (2017). How we transmit memories to other brains: constructing shared neural representations via communication. *Cereb. Cortex* 27, 4988–5000. doi: 10.1093/cercor/bhx202

Zatorre, R. J. (2001). Neural specializations for tonal processing. *Ann. N. Y. Acad. Sci.* 930, 193–210. doi: 10.1111/j.1749-6632.2001.tb05734.x

Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science* 342, 585–589. doi: 10.1126/science.1238414

Zatorre, R. J., Delhommeau, K., and Zarate, J. M. (2012). Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Front. Psychol.* 3:544. doi: 10.3389/fpsyg.2012.00544

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2018 Lumaca, Ravignani and Baggio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

MINI REVIEW
published: 16 May 2018
doi: 10.3389/fpsyg.2018.00762

# Music and the Meeting of Human Minds

*Alan R. Harvey**

*School of Human Sciences, The University of Western Australia, Perron Institute for Neurological and Translational Science, Perth, WA, Australia*

Over tens of thousands of years of human genetic and cultural evolution, many types and varieties of music and language have emerged; however, the fundamental components of each of these modes of communication seem to be common to all human cultures and social groups. In this brief review, rather than focusing on the development of different musical techniques and practices over time, the main issues addressed here concern: (i) when, and speculations as to why, modern *Homo sapiens* evolved musical behaviors, (ii) the evolutionary relationship between music and language, and (iii) why humans, perhaps unique among all living species, universally continue to possess two complementary but distinct communication streams. Did music exist before language, or vice versa, or was there a common precursor that in some way separated into two distinct yet still overlapping systems when cognitively modern *H. sapiens* evolved? A number of theories put forward to explain the origin and persistent universality of music are considered, but emphasis is given, supported by recent neuroimaging, physiological, and psychological findings, to the role that music can play in promoting trust, altruistic behavior, social bonding, and cooperation within groups of culturally compatible but not necessarily genetically related humans. It is argued that, early in our history, the unique socializing and harmonizing power of music acted as an essential counterweight to the new and evolving sense of self, to an emerging sense of individuality and mortality that was linked to the development of an advanced cognitive capacity and articulate language capability.

**Keywords: music, evolution, altruism, social cooperation, language, motherese, dopamine, oxytocin**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky*
*Independent Researcher,*
*Los Angeles, CA, United States*

***Reviewed by:***
*Guy Madison,*
*Umeå University, Sweden*
*Steven Robert Livingstone,*
*University of Wisconsin–River Falls,*
*United States*

***Correspondence:**
*Alan R. Harvey*
*alan.harvey@uwa.edu.au*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *15 January 2018*
***Accepted:*** *30 April 2018*
***Published:*** *16 May 2018*

***Citation:***
*Harvey AR (2018) Music and the Meeting of Human Minds. Front. Psychol. 9:762. doi: 10.3389/fpsyg.2018.00762*

In the twenty-first century, there are many different types and varieties of music, sung in different ways and played on many different types of musical instrument, and there are also many different types of language with different alphabets and modes of writing. Despite this biological, cultural, and behavioral diversity, the fundamental elements of music and language communication seem to be universals, present in all human cultures and social groups. There is an ever-increasing literature that focuses on the neuroscience, evolution, psychology, and sociobiology of music (e.g., Levitin, 2006; Patel, 2008; Morley, 2013; Harvey, 2017). In addressing the issue of "why music?", a number of key questions arise, including: (i) when did our species evolve musical behaviors, (ii) what is the evolutionary relationship between music and language, and (iii) why do humans universally continue to possess two complementary but distinct communication streams, streams that are processed by some common but also some distinct pathways and circuits within the human brain.

Language is the main way we learn and pass on skills, ideas, plans, and acquired knowledge within, and across, generations. It structures thought and permits intuitive reasoning, it is

expandable in content and allows the creation of virtual realities and the manipulation of symbols. But language, similar to music, also possesses emotional or prosodic content, involving changes in pitch, voice quality, loudness, and timing of delivery. Such changes convey information about the emotional state of the speaker and/or change the meaning of a word or phrase, and when vocalizing there are elements in common in the generation of phrases in music (singing) and speech (Patel, 2008). Interestingly then, studies using functional magnetic resonance imaging (fMRI) and other methods to monitor brain activity have revealed that elements of both language and music, including for example aspects of pitch and rhythm processing (Rogalsky et al., 2011; Perrachione et al., 2013), and perhaps even some elements of syntax (e.g., Sammler et al., 2013; Kunert et al., 2015, but see Bigand et al., 2014), are processed in closely related regions within the brain.

There are also, however, differences in brain circuitry. Spoken language, at least in right handers, is mostly left hemisphere biased, whereas music has a right hemisphere bias, although with training the networks involved with music processing generally show a leftward shift (Ellis et al., 2013; Thaut et al., 2014). Recent research has identified specific regions in the temporal lobe that appear to respond only to music or to speech (Angulo-Perkins et al., 2014; Norman-Haignere et al., 2015). In addition, music activates pathways and regions within the limbic system (Koelsch, 2014), a system associated with a number of functions including learning, memory, motivation, and emotional responsiveness. Music we appreciate and subjectively find arousing is associated with dopaminergic activity and activates reward centers such as the nucleus accumbens, located in the ventral striatum (Menon and Levitin, 2005; Salimpoor et al., 2011; Zatorre and Salimpoor, 2013; Mueller et al., 2015). The pathways that mediate interactions with other sensory modalities and links to motor systems controlling vocalization and body movement also show some left versus right bias for speech and music, respectively (Callan et al., 2006; Özdemir et al., 2006).

From an evolutionary perspective, the origins of modern language and articulate speech in *Homo sapiens,* and the extent to which language is encoded in the brain and/or influenced by cultural experience, remain controversial (e.g., Pinker, 1994; Hauser et al., 2002; Bickerton, 2007; Patel, 2008). Nonetheless, while obviously speculative, it seems likely that linguistic-related advances in neural hardware and software were superimposed on pre-existing neural capabilities, the latter including prosodic utterance as well as non-verbal behaviors such as hand movement (gesture), tool making, and facial expression (Mithen, 2005; Arbib et al., 2008; Aboitiz and Garcia, 2009; Fay et al., 2014). There was also likely a need for enhanced neural plasticity as well as improved attention mechanisms, greater capacity for abstract thought, a vastly improved speed of learning, and enhanced working memory and storage capability (Harvey, 2017). But if language was a key partner of our newly evolved cognitive power, and a primary driver of human cultural evolution, why did we also evolve, and why do we continue to use and enjoy, another communication stream, music? The presence of at least some shared linguistic and musical neural circuitry is not inconsistent with the proposal that, during the evolution of modern *H. sapiens*, language and music emerged from a common precursor possessed by our immediate ancestors, sometimes called a musilanguage (Brown, 2000) or protolanguage. But what does music do for us, as individuals and as a community, and where did the production and appreciation of music come from, and when? Can music have had meaningful evolutionary significance even though "it neither, plows, sows, weaves nor feeds" (Cross, 2001)?

In this brief review, the focus is on speculation about the earliest origins of music in *H. sapiens*, rather than discussing the subsequent genetic/cultural evolution of different musical genres, styles, and social practices: "Distinctions between the surface complexity of different musical styles and techniques do not tell us anything useful about the expressive purposes and power of music, or about the intellectual organization involved with its creation" (Blacking, 1976). Fossil evidence suggests that the earliest evidence of our species, *H. sapiens*, first appeared about 200,000 years ago, perhaps even as far back as 300,000 years ago (Hublin et al., 2017). Pieces of fossilized bone provide only limited information about the evolution of the modern human brain (Sherwood et al., 2008). More meaningful archeological evidence, strongly suggestive of relatively recent cognitive and behavioral advances in our species, comes from the discovery and analysis of cultural artifacts and tools (Henshilwood et al., 2002; Bouzouggar et al., 2007). Scientists can also trace lineage relationships in modern *H. sapiens*, to determine how genes, and especially the subtle regulation of those genes, have changed over the millennia. Coalescent modeling allows researchers to work backward from DNA samples from members of different human population groups in order to develop gene trees that identify the most likely common or shared ancestral population from which all living humans are derived. Lineage tracing based on changes to the Y chromosome genotype permits a study of paternal ancestry, and changes in the DNA within mitochondria allow an analysis of maternal lineage relationships (Underhill and Kivisild, 2007).

There are, of course, significant margins of error associated with such studies. Nonetheless the genetic data, when combined with evidence of cultural change, do suggest that the modern version of *H. sapiens*, the version with an advanced cognitive architecture and newly developed language capabilities, evolved within the last 100,000 years, perhaps even more recently. A study looking at the four major variants of African mitochondrial DNA has found that there was a rapid expansion of the group carrying of one of these haplotypes (L3) between 61,000 and 86,000 years ago, and that all humans living outside Africa share this phenotype (Atkinson et al., 2009; Soares et al., 2012). Within Africa, L3 carrying populations originally in Eastern Africa also increased, migrating into central and southern parts of the continent. Overall, while there were earlier migrations of *H. sapiens* out of Africa (Reyes-Centeno et al., 2014), it seems likely that the last major dispersal out of Africa by our founder population occurred somewhere between 60,000 to 70,000 years ago. Since then, human evolution has been driven by complex and dynamic interactions between genes, environment, and culture, yet both within and beyond Africa language and music remain universals, found in all living

human populations and cultures. For me, the most parsimonious explanation for the universality of both communication streams is that members of our founder population possessed both systems from the very beginning (Harvey, 2017). What is clear is that by at least 43,000 years ago our ancestors were already modifying bone to make flute-like instruments (Higham et al., 2012).

Numerous theories have been put forward in an attempt to explain the seeming universality of music, from its earliest evolutionary origins through to the present day, and the possible adaptive benefits that music brought, and continues to bring, to our species (Hauser and McDermott, 2003; Huron, 2003; Mithen, 2005; Harvey, 2017). These ideas are necessarily speculative, but scientific and psychological research over the past few decades does provide some useful clues and pointers. Music and its partner dance may play an important role in mate attraction and selection, it can help structure time, and it facilitates long-distance communication. Music also enhances cognitive and motor skill development, has mnemonic power, encourages group cooperation, and social interactions, helps to define cultural identity and perhaps even facilitates conflict resolution. Of course, there may have been multiple, and intersecting, advantages for our founder population to possess musical communication as part of its make-up. Here, I wish to emphasize the early evolutionary importance of music in the context of cooperative interactions and prosocial behavior in *H. sapiens*.

One intriguing idea is that music-like capabilities may have evolved to enable mothers to interact with infants during the first few months after birth, encouraging parent-infant attachment before babies learn to talk and use language (DeCasper and Fifer, 1980; Trehub, 2000; Falk, 2004; Levitin, 2006; Trevarthan, 2008). This infant-directed communication, important in driving the behavioral, perceptual, cognitive, emotional, and social development of children, is generally known as motherese. Motherese has particular characteristics; it is rhythmic, usually delivered slowly and at higher pitch with exaggerated pitch changes that often slide from one to another, it is calming and reassuring, and is associated with other key social and emotional interactions that involve gesture, eye contact, and facial expression. However, while motherese may sound musical, the scientific evidence currently available does not, in my view, clearly establish if it is "protomusical" or "protolinguistic" in nature (Harvey, 2017). This prosodic maternal-infant communication system may in fact be more akin to the more "primitive" system used by our ancestors, prior to the more recent evolution of the two communication streams of articulate speech and music.

In adult life, speech usually involves turn taking, one person talks and another listens (Levinson, 2016), whereas music is capable of integrating and coordinating feelings and emotions within even a large group. Music, and its ally dance, require both energy and time commitment, so from an evolutionary perspective: "..if music had no value whatsoever, one might expect strong selection against musical behavior. Recent evidence from subjects with congenital amusia indicates that the necessary genetic variance is present in human populations, albeit at a very low frequency. So why have not quiet, better-rested non-musical humans out-reproduced and replaced their musical conspecifics?" (Fitch, 2006).

A possible reason for the continued universality of music, as argued in detail elsewhere (Harvey, 2017), is that the evolution of language and the modern human mind, and a new capacity for transgenerational communication and mental time travel (Suddendorf and Corballis, 2007), led to a more sharply defined sense of individuality. Our species became, as others have described it, a "society of selves" (Humphrey, 2008), with greater levels of social complexity and uncertainty, and increasingly aware of our own mortality and finite life span: "Only human beings guide their behavior by a knowledge of what happened before they were born and a preconception of what may happen after they are dead; thus only humans find their way by a light that illuminates more than the patch of ground they stand on" (Medawar, 1977). The relatively indeterminate nature of music allowed humans to interact, infer, and actively share experiences even though at the individual level each person may have been experiencing different emotions and thoughts: "Music allows participants to explore the prospective consequences of their actions and attitudes toward others within a temporal framework that promotes the alignment of participants' sense of goals" (Cross, 2009). In other words the rewarding, harmonizing power of music and music-related activities such as dance may have acted as important and essential counterweights to the individualization experienced by increasingly intelligent, articulate, and apprehensive members of our species. As Steven Brown wrote: ". . .the straightforward evolutionary implication is that human musical capacity evolved because groups of musical hominids out-survived groups of non-musical hominids due to a host of factors related to group-level cooperation and coordination" (Brown, 2000).

If music did indeed play a key evolutionary role in facilitating and rewarding social bonding and mutual cooperation in *H. sapiens*, is there any evidence of this relationship at the neurological level? Neuroimaging studies have revealed that positive, optimistic thoughts activate specific regions such as the anterior cingulate cortex, the amygdala and parts of the prefrontal cortex (Sharot et al., 2007), similar to some of the regions associated with positive emotional responses to music (e.g., Salimpoor et al., 2009; Zatorre and Salimpoor, 2013; Koelsch, 2014). Altruistic acts are a form of social behavior that result in some form of benefit for the receiver and usually, but not always, at least some negative consequences for the giver. In humans, reciprocal altruism, involving mutual cooperative advantage (Trivers, 1971; Fehr and Fischbacher, 2003), has a beneficial impact not only on the viability and welfare of individuals but can also lead to better cooperation with kin as well as promote pro-social attitudes within larger groups of culturally compatible but unrelated individuals (Boyd and Richerson, 2009; Kurzban et al., 2015). Significantly, when subjects are performing mutually cooperative tasks that are rewarding and require empathy and an appreciation of social context and another person's feelings, there is increased activity in areas such as the nucleus accumbens, amygdala, anterior cingulate cortex,

superior temporal cortex, temporoparietal junction, and several specific regions in prefrontal cortex (Rilling et al., 2002, 2008; O'Doherty, 2004). Once again, these regions overlap many of the regions that are active when listening to familiar and emotionally rewarding music, music that can reinforce the sharing of affective states.

There is also neurochemical and physiological evidence to support the idea that music acts to promote empathy and cooperative behavior. Music alters the expression of a number of agents that modulate brain activity and that are associated with reward and a sense of well-being, including not only neurotransmitters such as dopamine but also several hormones (Chanda and Levitin, 2013; Zatorre and Salimpoor, 2013). Pain sensitivity, thought to be a surrogate for altered endogenous opioid levels, is lowered during participation in group music activities, including choral singing and dance (Dunbar et al., 2012; Tarr et al., 2015), and music can reduce levels of the stress hormone cortisol (Khalfa et al., 2003).

In humans the hormone oxytocin has been shown to promote the expression of positive emotions and prosocial altruistic attitudes (Hurlemann et al., 2010; Marsh et al., 2015; Leppanen et al., 2017). When applied as a spray via the nose, oxytocin reduces anxiety and fosters trust and emotional empathy (Kosfeld et al., 2005; Baumgartner et al., 2008), and genetic variants of the oxytocin receptor are linked to differences in the way humans empathize and engage in prosocial activities (Rodrigues et al., 2009; Aspé-Sánchez et al., 2016). In a study that also analyzed stress and cortisol levels in participants, it was recently reported that oxytocin levels measured in saliva are *decreased* during choral but not solo singing (Schladt et al., 2017). However, measurement of salivary oxytocin levels do not necessarily reflect levels within the specific regions of the brain that possess oxytocin receptors (Mitre et al., 2016; Lin et al., 2018), and there is a clearer and more consistent relationship between blood plasma levels of oxytocin and levels found in cerebrospinal fluid (Valstad et al., 2017). Interestingly then, circulating levels of the hormone oxytocin were found to be increased only under conditions in which singers were asked to *improvise* together (Keeler et al., 2015). Musical improvisation, presumably comparable to the type of activity experienced by early humans, requires that participants innovate, engage, and socially interact with other members of the group. Consistent with this, compared to other creative group activities, singing encourages more rapid inter-individual bonding, apparently short-circuiting the usual way that social relationships are established: "The capacity of singing to bond groups of relative strangers in humans may have played a crucial role in allowing modern humans to create and maintain much larger social networks than their evolutionary relatives. . ." (Pearce et al., 2015).

For many thousands of years, the evolution and development of musical styles and practices, as well as individual musical abilities, has been driven by a complex interplay between genes and culture. Evidence for the vertical transmission of at least some aspects of musical behavior in human populations comes from studies showing that genetic variants in a number of neuromodulator receptors are linked to musical memory, aptitude, and creativity (Liu et al., 2016; Oikkonen et al., 2016; Mariath et al., 2017). Furthermore, lineage studies have revealed a correlation between variation in folk musical styles and genetic variance within populations (Pamjav et al., 2012; Brown et al., 2013; Le Bomin et al., 2016), all lending support to the proposal that musical traits are heritable.

In summary, it is argued that music and music-related behaviors, operating alongside language and articulate speech, were important in helping to promote emotional synergy, social bonding, and foster group-level cooperation and coordination early in human evolution. As suggested elsewhere (Cross, 2003), music may have been especially suited for this purpose because it is usually risk free and of indeterminate meaning. In a group context, music allows humans to interact and share experiences during a particular musical activity even though each person may have very different outlooks and goals, and be involved in a range of real or imagined relationships. "Wherever humans live, and however, they have organized their societies, they exhibit a behavioral peculiarity of gathering from time to time to sing and dance together in a group. By featuring both human song and entrainment (in the dancing movements and perhaps clapping performed in synchrony with the singing/music), such behavior qualifies as human music. Indeed, the fact that it occurs in every human culture, and indeed subculture, without exception, unless deliberately suppressed by severe sanctions against it, marks this phenomenon as the most universal human behavior of a musical kind on record" (Merker et al., 2015).

An appreciation of the evolutionary significance of music, and the impact that the universal of music has had in promoting social harmony and an individual's sense of well-being in our species, can only help when advocating for the importance of music in education and for therapeutic purposes in the clinic (Harvey, 2012). Music and associated synchronized behaviors remain core human attributes, promoters of prosocial, empathic behaviors, and cultural cohesion, and it is to be hoped that humans will continue to harness music's power, for the good of society.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

## REFERENCES

Aboitiz, F., and Garcia, R. (2009). Merging of phonological and gestural circuits in early language evolution. *Rev. Neurosci.* 20, 71–84. doi: 10.1515/REVNEURO.2009.20.1.71

Angulo-Perkins, A., Aubé, W., Peretz, I., Barios, F. A., Armony, J. L., and Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: differences between musicians and non-musicians. *Cortex* 59, 126–137. doi: 10.1016/j.cortex.2014.07.013

Arbib, M. A., Liebal, K., and Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Curr. Anthrop.* 49, 1053–1063. doi: 10.1086/593015

Aspé-Sánchez, M., Moreno, M., Rivera, M. I., Rossi, A., and Ewer, J. (2016). Oxytocin and vasopressin receptor gene polymorphisms: role in

social and psychiatric traits. *Front. Neurosci.* 9:510. doi: 10.3389/fnins.2015.00510

Atkinson, Q. D., Gray, R. D., and Drummond, A. J. (2009). Bayesian coalescent interference of major human mitochondrial DNA haplotype expansions in Africa. *Proc. Biol. Sci.* 276, 367–373. doi: 10.1098/rspb.2008.0785

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., and Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and adaptation in humans. *Neuron* 58, 639–650. doi: 10.1016/j.neuron.2008.04.009

Bickerton, D. (2007). Language evolution: a brief guide for linguists. *Lingua* 117, 510–526. doi: 10.1016/j.lingua.2005.02.006

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., and Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations revel the contribution of auditory short-term memory. *Front. Syst. Neurosci.* 8:94. doi: 10.3389/fnsys.2014.00094

Blacking, J. (1976). *How Musical is Man?* London: Faber and Faber.

Bouzouggar, A., Barton, N., Vanhaeren, M., d'Errico, F., Collcutt, S., Higham, T., et al. (2007). 82,000-year-old shell beads from North Africa and implications for the origins of modern human behavior. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9964–9969. doi: 10.1073/pnas.0703877104

Boyd, R., and Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 3281–3288. doi: 10.1098/rstb.2009.0134

Brown, S. (2000). "The "musilanguage" model of human evolution," in *Origins of Music*, eds N. Wallin, B. Merker and S. Brown (Cambridge MA: MIT Press), 271–300.

Brown, S., Savage, P. E., Ko, A. M., Stoneking, M., Ko, Y. C., Loo, J. H., et al. (2013). Correlations in the population structure of music, genes and language. *Proc. Biol. Soc.* 281:20132072. doi: 10.1098/rspb.2013.2072

Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., et al. (2006). Song and speech: brain regions involved with perception and covert production. *Neuroimage* 31, 1327–1342. doi: 10.1016/j.neuroimage.2006.01.036

Chanda, M. L., and Levitin, D. J. (2013). The neurochemistry of music. *Trends Cogn. Sci.* 17, 179–193. doi: 10.1016/j.tics.2013.02.007

Cross, I. (2001). Music, mind and evolution. *Psych. Music* 29, 95–112. doi: 10.1177/0305735601291007

Cross, I. (2003). "Music, cognition, culture and evolution," in *The Cognitive Neuroscience of Music*, eds I. Peretz and R. Zatorre (Oxford: Oxford University Press), 42–56. doi: 10.1093/acprof:oso/9780198525202.003.0004

Cross, I. (2009). The evolutionary nature of musical meaning. *Music. Sci.* 179–200. doi: 10.3389/fpsyg.2017.00494

DeCasper, A. J., and Fifer, W. P. (1980). Of human bonding: newborns prefer their motheers' voices. *Science* 208, 1174–1176. doi: 10.1126/science.7375928

Dunbar, R. I. M., Kaskatis, K., MacDonald, I., and Barra, V. (2012). Performance of music elevates pain threshold and positive affect: implications for the evolutionary function of music. *Evol. Psychol.* 10, 688–702. doi: 10.1177/147470491201000403

Ellis, R. J., Bruijn, B., Norton, A. C., Winner, E., and Schlaug, G. (2013). Training-mediated leftward asymmetries during music processing: a cross-sectional and longitudinal fMRI analysis. *Neuroimage* 75, 97–107. doi: 10.1016/j.neuroimage.2013.02.045

Falk, D. (2004). Prelinguistic evolution in early hominids: whence motherese? *Behav. Brain Sci.* 27, 491–503.

Fay, N., Lister, C. J., Ellison, T. M., and Goldin-Meadow, S. (2014). Creating a communication system from scratch: gesture beats vocalization hands down. *Front. Psychol.* 5:354. doi: 10.3389/fpsyg.2014.00354

Fehr, E., and Fischbacher, U. (2003). The nature of human altruism. *Nature* 425, 785–791. doi: 10.1038/nature02043

Fitch, T. W. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009

Harvey, A. R. (2012). "Evolution, music and neurotherapy," in *Pragmatic Evolution: Applications of Evolutionary Theory*, ed. A. Poiani (Cambridge: Cambridge University Press), 150–163.

Harvey, A. R. (2017). *Music, Evolution, and the Harmony of Souls*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198786856.001.0001

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579.

Hauser, M. D., and McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nat. Neurosci.* 6, 663–668. doi: 10.1038/nn1080

Henshilwood, C. S., d'Errico, F., Yates, R., Jacobs, Z., Tribolo, C., Duller, G. A., et al. (2002). Emergence of modern human behaviour: Middle Stone Age engravings from South Africa. *Science* 295, 1278–1280. doi: 10.1126/science.1067575

Higham, T., Basell, L., Jacobi, R., Wood, R., Bronk Ramsey, C., and Conard, N. J. (2012). Testing models for the beginnings of the Aurignacian and the advent of figurative art and music: the radiocarbon chronology of Geißenklösterle. *J. Hum. Evol.* 62, 664–676. doi: 10.1016/j.jhevol.2012.03.003

Hublin, J. J., Ben-Ncer, A., Bailey, S. E., Freidline, S. E., Neubauer, S., Skinner, M. M., et al. (2017). New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature* 546, 293–296. doi: 10.1038/nature22335

Humphrey, N. (2008). The society of selves. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 745–754. doi: 10.1098/rstb.2006.2007

Hurlemann, R., Patin, A., Onur, O. A., Cohen, M. X., Baumgartner, T., Metzler, S., et al. (2010). Oxytocin enhances amygdala-dependent, socially reinforced learning and emotional empathy in humans. *J. Neurosci.* 30, 4999–5007. doi: 10.1523/JNEUROSCI.5538-09.2010

Huron, D. (2003). "Is music an evolutionary adaptation?," in *The Cognitive Neuroscience of Music*, eds I. Peretz and R. L. Zatorre (Oxford: Oxford University Press).

Keeler, J. R., Roth, E. A., Neuser, B. L., Spitsbergen, J. M., Waters, D. J., and Vianney, J. M. (2015). The neurochemistry and social flow of singing: bonding and oxytocin. *Front. Hum. Neurosci.* 9:518. doi: 10.3389/fnhum.2015.00518

Khalfa, S., Bella, S. D., Roy, M., Peretz, I., and Lupien, S. J. (2003). Effects of relaxing music on salivary cortisol level after psychological stress. *Ann. N. Y. Acad. Sci.* 999, 374–376. doi: 10.1196/annals.1284.045

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., and Fehr, E. (2005). Oxytocin increases trust in humans. *Nature* 435, 673–676. doi: 10.1038/nature03701

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., and Hagoort, P. (2015). Music and language syntax interact in Broca's area: an fMRI study. *PLoS One* 10:e0141069. doi: 10.1371/journal.pone.0141069

Kurzban, R., Burton-Chellew, M. N., and West, S. A. (2015). The evolution of altruism in humans. *Annu. Rev. Psychol.* 66, 575–599. doi: 10.1146/annurev-psych-010814-015355

Le Bomin, S., Lecointre, G., and Heyer, E. (2016). The evolution of musical diversity: the key role of vertical transmission. *PLoS One* 11:e0151570. doi: 10.1371/journal.pone.0151570

Leppanen, J., Ng, K. W., Tchanturia, K., and Treasure, J. (2017). Meta-analysis of the effects of intranasal oxytocin on interpretation and expression of emotions. *Neurosci. Biobehav. Rev.* 78, 125–144. doi: 10.1016/j.neubiorev.2017.04.010

Levinson, S. C. (2016). Turn-taking in human communication – origins and implications for language processing. *Trends Cogn. Sci.* 20, 6–14. doi: 10.1016/j.tics.2015.10.010

Levitin, D. J. (2006). *This Is Your Brain on Music*. New York, NY: Dutton.

Lin, Y.-T., Hsieh, T.-Y., Tsai, T.-C., Chen, C.-C., Huang, C.-C., and Hsu, K.-S. (2018). Conditional deletion of hippocampal CA2/CA3a oxytocin receptors impairs the persistence of long-term social recognition memory in mice. *J. Neurosci.* 38, 1218–1231. doi: 10.1523/JNEUROSCI.1896-17.2017

Liu, X., Kanduri, C., Oikkonen, J., Karma, K., Raijas, P., Ukkola-Vuoti, L., et al. (2016). Detecting signatures of positive selection associated with musical aptitude in the human genome. *Sci. Rep.* 6:21198. doi: 10.1038/srep21198

Mariath, L. M., Silva, A. M. D., Kowalski, T. W., Gattino, G. S., Araujo, G. A., Figueiredo, F. G., et al. (2017). Music genetics research: association with musicality of a polymorphism in the AVPRA gene. *Genet. Mol. Biol.* 40, 421–429. doi: 10.1590/1678-4685-GMB-2016-0021

Marsh, N., Scheele, D., Gerhardt, H., Strang, S., Enax, L., Weber, B., et al. (2015). The neuropeptide oxytocin induces a social altruism bias. *J. Neurosci.* 35, 15696–15701. doi: 10.1523/JNEUROSCI.3199-15.2015

Medawar, P. B. (1977). *The Life Science: Current Ideas in Biology*. London: Wildwood House.

Menon, V., and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage* 28, 175–184. doi: 10.1016/j.neuroimage.2005.05.053

Merker, B., Morley, I., and Zuidema, W. (2015). Five fundamental constraints on theories of the origins of music. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370:20140095. doi: 10.1098/rstb.2014.0095

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Orion.

Mitre, M., Marlin, B. J., Schiavo, J. K., Morina, E., Norden, S. E., Hackett, T. A., et al. (2016). A distributed network for social cognition enriched for oxytocin receptors. *J. Neurosci.* 36, 2517–2535. doi: 10.1523/JNEUROSCI.2409-15.2016

Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology and the Origins of Musicality*. Oxford: Oxford University Press. doi: 10.1093/acprof:osobl/9780199234080.001.0001

Mueller, K., Fritz, T., Mildner, T., Richter, M., Schulze, K., Lepsien, J., et al. (2015). Investigating the dynamics of the brain response to music: a central role of the ventral striatum/nucleus accumbens. *Neuroimage* 116, 68–79. doi: 10.1016/j.neuroimage.2015.05.006

Norman-Haignere, S., Kanwisher, N. G., and McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296. doi: 10.1016/j.neuron.2015.11.035

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* 14, 769–776. doi: 10.1016/j.conb.2004.10.016

Oikkonen, J., Onkamo, P., Järvelä, I., and Kanduri, C. (2016). Convergent evidence for the molecular basis of musical traits. *Sci. Rep.* 6:39707. doi: 10.1038/srep39707

Özdemir, E., Norton, A., and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage* 33, 628–635. doi: 10.1016/j.neuroimage.2006.07.013

Pamjav, H., Juhász, Z., Zalán, A., Németh, E., and Damdin, B. (2012). A comparative phylogenetic study of genetics and folk music. *Mol. Genet. Genomics* 287, 337–349. doi: 10.1007/s00438-012-0683-y

Patel, A. D. (2008). *Music, Language and the Brain*. New York: Oxford University Press.

Pearce, E., Launay, J., and Dunbar, R. I. M. (2015). The ice-breaker effect: singing mediates fast social bonding. *R. Soc. Open Sci* 2:150221. doi: 10.1098/rsos.150221

Perrachione, T. K., Fedorenko, E. G., Vinke, L., Gibson, E., and Dilley, L. C. (2013). Evidence for shared cognitive processing of pitch in music and language. *PLoS One* 8:e73372. doi: 10.1371/journal.pone.0073372

Pinker, S. (1994). *The Language Instinct*. London: Penguin. doi: 10.1037/e412952005-009

Reyes-Centeno, H., Ghirotto, S., Détroit, F., Grimaud-Hervé, D., Barbujani, A. G., and Harvati, K. (2014). Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7248–7253. doi: 10.1073/pnas.1323666111

Rilling, J. K., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. *Neuron* 67, 511–521. doi: 10.1016/S0896-6273(02)00755-9

Rilling, J. K., King-Cass, B., and Sanfey, A. G. (2008). The neurobiology of social decision-making. *Curr. Opin. Neurobiol.* 18, 159–165. doi: 10.1016/j.conb.2008.06.003

Rodrigues, S. M., Saslow, L. R., Garcia, N., John, O. P., and Keltner, D. (2009). Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21437–21441. doi: 10.1073/pnas.0909579106

Rogalsky, C., Rong, F., Saberi, K., and Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* 31, 3843–3852. doi: 10.1523/JNEUROSCI.4515-10.2011

Salimpoor, V. N., Benovoy, M., Larcher, K., Dagher, A., and Zatorre, R. J. (2011). Anatomically distinct dopamine release during anticipation and experience of peak emotion in music. *Nat. Neurosci.* 14, 257–264. doi: 10.1038/nn.2726

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., and Zatorre, R. J. (2009). The rewarding aspects of music listening are related to the degree of emotional arousal. *PLoS One* 4:e7487. doi: 10.1371/journal.pone.0007487

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H. J., et al. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *Neuroimage* 64, 134–146. doi: 10.1016/j.neuroimage.2012.09.035

Schladt, T. M., Nordmann, G. C., Emilius, R., Kudielka, B. M., de Jong, T. R., and Neumann, I. D. (2017). Choir versus solo singing: effects on mood, and salivary oxytocin and cortisol concentrations. *Front. Hum. Neurosci.* 11:430. doi: 10.3389/fnhum.2017.00430

Sharot, T., Riccardi, A. M., Raio, C. M., and Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature* 450, 102–105. doi: 10.1038/nature06280

Sherwood, C. C., Subiaul, F., and Zawidzki, T. W. (2008). A natural history of the human mind: tracing evolutionary changes in brain and cognition. *J. Anat.* 212, 426–454. doi: 10.1111/j.1469-7580.2008.00868.x

Soares, P., Alshamali, F., Pereira, J. B., Fernandes, V., Silva, N. M., Afonso, C., et al. (2012). The expansion of mtDNA haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* 29, 915–927. doi: 10.1093/molbev/msr245

Suddendorf, Y., and Corballis, M. C. (2007). The evolution of foresight: what is mental time travel, and is it unique to humans? *Behav. Brain Sci.* 30, 299–351. doi: 10.1017/S0140525X07001975

Tarr, B., Launay, J., Cohen, E., and Dunbar, R. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biol. Lett.* 11:20150767. doi: 10.1098/rsbl.2015.0767

Thaut, M. H., Trimarchi, P. D., and Parsons, L. M. (2014). Human brain basis of musical rhythm perception: common and distinct neural substrates for meter, tempo, and pattern. *Brain Sci.* 4, 428–452. doi: 10.3390/brainsci4020428

Trehub, S. (2000). "Human processing dispositions and musical universals," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 427–448.

Trevarthan, C. (2008). The musical art of infant conversation: narrating in the time of sympathetic experience, without rational interpretation, before words. *Music. Sci.* 12(Suppl.), 15–46. doi: 10.1177/1029864908012001021

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57. doi: 10.1086/406755

Underhill, P. A., and Kivisild, T. (2007). Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annu. Rev. Genet.* 41, 539–564. doi: 10.1146/annurev.genet.41.110306.130407

Valstad, M., Alvares, G. A., Egknud, M., Matziorinis, A. M., Andreassen, O. A., Westlye, L. T., et al. (2017). The correlation between central and peripheral oxytocin concentrations: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* 78, 117–124. doi: 10.1016/j.neubiorev.2017.04.017

Zatorre, R. J., and Salimpoor, V. N. (2013). From perception to pleasure: music and its neural substrates. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10430–10437. doi: 10.1073/pnas.1301228110

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2018 Harvey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 16 July 2018
doi: 10.3389/fnins.2018.00475

# Rhesus Monkeys (*Macaca mulatta*) Sense Isochrony in Rhythm, but Not the Beat: Additional Support for the Gradual Audiomotor Evolution Hypothesis

*Henkjan Honing*[1]*, *Fleur L. Bouwer*[1], *Luis Prado*[2] and *Hugo Merchant*[2]

[1] Amsterdam Brain and Cognition, Institute for Advanced Study, Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, Netherlands, [2] Department of Cognitive Neuroscience, Instituto de Neurobiología, Universidad Nacional Autonoma de México, Santiago de Querétaro, Mexico

Charles Darwin suggested the perception of rhythm to be common to all animals. While only recently experimental research is finding some support for this claim, there are also aspects of rhythm cognition that appear to be species-specific, such as the capability to perceive a regular pulse (or beat) in a varying rhythm. In the current study, using EEG, we adapted an auditory oddball paradigm that allows for disentangling the contributions of beat perception and isochrony to the temporal predictability of the stimulus. We presented two rhesus monkeys (*Macaca mulatta*) with a rhythmic sequence in two versions: an isochronous version, that was acoustically accented such that it could induce a duple meter (like a march), and a jittered version using the same acoustically accented sequence but that was presented in a randomly timed fashion, as such disabling beat induction. The results reveal that monkeys are sensitive to the isochrony of the stimulus, but not its metrical structure. The MMN was influenced by the isochrony of the stimulus, resulting in a larger MMN in the isochronous as opposed to the jittered condition. However, the MMN for both monkeys showed no interaction between metrical position and isochrony. So, while the monkey brain appears to be sensitive to the isochrony of the stimulus, we find no evidence in support of beat perception. We discuss these results in the context of the gradual audiomotor evolution (GAE) hypothesis (Merchant and Honing, 2014) that suggests beat-based timing to be omnipresent in humans but only weakly so or absent in non-human primates.

**Keywords: music, rhythm, beat perception, ERP, MMN**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky,*
*Braavo! Enterprises, United States*

***Reviewed by:***
*Christopher I. Petkov,*
*Newcastle University, United Kingdom*
*Aniruddh Patel,*
*Tufts University, United States*

***Correspondence:**
*Henkjan Honing*
*honing@uva.nl*

***Specialty section:***
*This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience*

***Received:*** *30 September 2017*
***Accepted:*** *22 June 2018*
***Published:*** *16 July 2018*

***Citation:***
*Honing H, Bouwer FL, Prado L and Merchant H (2018) Rhesus Monkeys (Macaca mulatta) Sense Isochrony in Rhythm, but Not the Beat: Additional Support for the Gradual Audiomotor Evolution Hypothesis. Front. Neurosci. 12:475. doi: 10.3389/fnins.2018.00475*

## INTRODUCTION

The interest in rhythm cognition in non-human animals is motivated by the search for signs of *musicality* as a means to get an insight in the evolutionary and causal processes underlying human musicality (Trehub, 2003; Honing and Ploeger, 2012; Hoeschele et al., 2015; Honing et al., 2015; Trehub et al., 2015; Honing, 2018b)[1]. Most animals show at least some sort of rhythmic behavior,

[1] *Musicality*, in all its complexity, can be defined as a natural, spontaneously developing set of traits [designed for the perception and production of music] based on and constrained by our cognitive abilities and their underlying biology. As such, *music*, in all its variety, can be defined as a social, and cultural construct based on that very musicality (Honing, 2018a).

like walking, flying, crawling, or swimming. It is hence not unnatural to think that the perception (and enjoyment) of rhythm might be shared by most animals, as was argued by Darwin (1871) and Patel (2014). While only recently experimental research is finding some support for this claim (Wilson and Cook, 2016), there are also aspects of rhythm cognition that appear to be species-specific (Fitch, 2013), such as the capability to perceive a regular pulse in a varying rhythm (i.e., one level of a metrical structure) and consequently being able to synchronize to it (i.e., rhythmic entrainment), referred to as *beat-based timing* (Merchant and Honing, 2014).

Beat-based timing in humans is a complex neurocognitive phenomenon that depends on a dynamic interaction between auditory and motor systems in the brain (Grahn and Brett, 2009; Morillon et al., 2014; Patel and Iversen, 2014; Hoeschele et al., 2015; Merchant and Yarrow, 2016; Ross et al., 2017). It is hypothesized to be facilitated by bidirectional, and potentially causal links between the auditory and motor areas in the brain, including the motor cortico-basal-ganglia-thalamo-cortical (mCBGT) circuit, that appear to be more developed in humans as opposed to non-human primates and related species (Patel et al., 2009; Mendoza and Merchant, 2014; Patel and Iversen, 2014; Petkov and Jarvis, 2014; Merchant et al., 2015a; Wilson and Cook, 2016).

These observations lead to the *gradual audiomotor evolution (GAE)* hypothesis (Merchant and Honing, 2014) that suggests beat-based timing to be gradually developed in primates, peaking in humans but present only with limited properties in other non-human primates, while humans share interval-based timing with all non-human primates and related species. Thus, the GAE hypothesis accommodates the fact that the performance of rhesus monkeys is comparable to humans in single interval tasks—such as categorization, interval reproduction, and interception—, but differs in multiple interval tasks, such as synchronization, continuation, and rhythmic entrainment (Honing and Merchant, 2014; Merchant and Honing, 2014).

In the current paper, we will focus on beat and isochrony perception as two key components of musicality (Merchant et al., 2015b; Honing, 2018a), and provide further evidence for the GAE hypothesis by studying rhythm perception in two rhesus monkeys (*Macaca mulatta*). For this study we used an existing ERP paradigm that allows for testing and disentangling the contributions of beat perception and isochrony to the temporal predictability of the stimulus (Bouwer et al., 2016).

## EARLIER WORK

Most existing animal studies on beat-based-timing and rhythmic entrainment (Wilson and Cook, 2016) have used behavioral methods to probe the presence of beat perception, such as tapping tasks (Zarco et al., 2009; Hasegawa et al., 2011; Hattori et al., 2015) or measuring head bobs (Patel et al., 2009; Schachner et al., 2009; Cook et al., 2013). However, if the production of synchronized movement to sound or music is not observed in certain species, this is no evidence for the absence of beat perception. It could well be that while certain species are not able to synchronize their movements to a regular beat, they may be capable of *beat perception* (i.e., the ability to perceive a regular pulse in a temporally and/or acoustically varying rhythm; Honing, 2012). With behavioral methods that rely on overt motoric responses it is difficult to separate between the contribution of perception and action. More direct, electrophysiological measures such as auditory event-related brain potentials (ERPs) allow to test for neural correlates of rhythm cognition, including beat perception (Honing et al., 2014).

While the vast majority of previous studies on animals have used implanted electrodes to record electroencephalograms (EEG) (Javitt et al., 1994; Laughlin et al., 1999; Pincze et al., 2001), non-invasive electrophysiological techniques such as scalp recorded evoked potentials (EP) and event-related potentials (ERP) are considered an attractive alternative. Next to being a mandatory requirement for studying some non-human primates such as chimpanzees (Fukushima et al., 2010; Hirata et al., 2013), these methods allow for a direct comparison between human and non-human primates. As such they have contributed to establishing animal models of the human brain and human brain disorders (Godlove et al., 2011; Gil-da-Costa et al., 2013), a better understanding of the neural mechanisms underlying the generation of human evoked EP/ERP components (Fishman and Steinschneider, 2012), as well as delineating cross-species commonalities and differences in brain functions, including rhythm cognition (Ueno et al., 2008, 2010; Fukushima et al., 2010; Reinhart et al., 2012; Hirata et al., 2013; Itoh et al., 2015). We will describe the most relevant ERP components for rhythm perception below.

## USING ERPS IN MEASURING BEAT PERCEPTION

The mismatch negativity (MMN) is an auditory event-related component that was shown to be sensitive to rhythmic violations in both humans and monkeys (see Honing et al., 2014 for a review). The MMN can be used as an index of a violation of temporal expectation using an oddball paradigm, by identifying a negative peak shortly after the deviant (the "oddball") that is maximal at fronto-central midline electrode sites and has sources in the auditory cortices and in the inferior frontal gyrus (i.e., not primarily reflecting motor cortex activity; Gil-da-Costa et al., 2013). The larger the violation of rhythmic expectations, the larger is the amplitude of the MMN (Näätänen et al., 2007; Winkler, 2007). The MMN has been shown to be indicative of beat perception in humans, with deviants on the beat within a repeating metrical auditory pattern eliciting a larger MMN than deviants off the beat (Ladinig et al., 2009, 2011; Winkler et al., 2009; Bouwer et al., 2014, 2016; Honing et al., 2014; Bouwer and Honing, 2015; Mathias et al., 2016).

The P3a, thought to reflect the redirection of attention to a deviant stimulus (Polich, 2007) and possibly an index of the conscious perception of a deviant (cf. Mathias et al., 2016; Peretz, 2016), often emerges just after the MMN, and has a latency of 200–250 ms in humans and between 100 and 250 ms in rhesus

monkeys (Picton et al., 1974; Polich, 2007). Gil-da-Costa et al. (2013) provided functional evidence that the neural generators of both MMN and P3a may be homologous in humans and monkeys, despite the observed differences in latency (see **Table 1** for an overview).

In addition to the ERP responses that reflect the detection of a deviant stimulus, the P1 and N1 responses, two early auditory event-related components, have been shown to be sensitive to the timing of the stimulus presentation (Costa-Faidella et al., 2011; Schwartze et al., 2013; Pereira et al., 2014; Teichert, 2016). In general, both P1 and N1 have an inverse relationship of ERP amplitude with temporal predictability (Javitt et al., 2000; Costa-Faidella et al., 2011; Schwartze et al., 2013). As such they are not indicative of beat perception *per se*, but of the temporal predictability of the stimulus. For instance, Schwartze et al. (2013) showed the P1 and N1 to be smaller in an isochronous as opposed to a jittered rhythmic sequence. Generally, increasing the predictability of the auditory stimulation (both in stimulus timing and stimulus probability) leads to a pronounced N1 attenuation (Costa-Faidella et al., 2011). Both components are maximal over fronto-central electrodes and there is some consensus on their homolog in rhesus monkeys, most notably on the N1 (Teichert, 2016) (see **Table 1** for an overview).

## RHYTHM COGNITION IN MONKEYS

Recently, Honing et al. (2012) were able to show, for the first time, that an MMN-like response can be measured in rhesus monkeys (*M. mulatta*). [See (Ueno et al., 2008) using a similar method in a chimpanzee (*Pan troglodytes),* and (Gil-da-Costa et al., 2013) for a recent study comparing humans and macaques (*Macaca fascicularis*)].

In addition, Honing et al. (2012) showed a sensitivity of the MMN in response to pitch deviants and infrequent omissions, showing that it was possible, in principle, to use an identical paradigm for human and non-human primates to probe beat perception. However, and contrary to what was found in human adults and infants (Winkler et al., 2009; Bouwer et al., 2014), no difference was found in the MMN in response to omissions in beat and offbeat positions. This lead to the conclusion that rhesus monkeys are unable to sense the beat (Honing et al., 2012). In addition, a strong response was found for onsets of rhythmic groups suggesting a sensitivity to rhythmic structure (A similar result was reported in Selezneva et al., 2013 showing large responses to changes in a repeating temporal pattern, while measuring gaze and facial expressions in monkeys).

However, Bouwer et al. (2014) pointed out that the earlier paradigm (Winkler et al., 2009; Honing et al., 2012) needs additional controls to be certain that any effects (or the lack thereof) are due to beat perception, and not, for instance, a result of pattern matching, acoustic variability or sequential learning. While rhesus monkeys have apparently little or no ability to perceive a beat, they are able to detect the regularity of an isochronous visual or auditory metronome (Zarco et al., 2009; Merchant et al., 2015b; Gámez et al., 2018). This suggests a capacity for making temporal predictions which most likely depends on absolute interval perception (Merchant and Honing, 2014). As such monkeys might not have beat perception, but could still be able to sense the regularity in an isochronous stimulus.

To examine the perception of isochrony, several studies have compared the responses to temporally regular, isochronous sequences with the responses to temporally irregular, jittered sequences (Schwartze et al., 2011; Teki et al., 2011; Fujioka et al., 2012). The prediction of events in jittered sequences has been suggested to rely on absolute interval perception, while the prediction of events in isochronous sequences has been suggested to be based on beat perception (Schwartze et al., 2011; Fujioka et al., 2012). However, it is possible to predict events in isochronous sequences on the basis of absolute interval perception alone, which may explain why macaques, with little or no ability to perceive a beat (Honing et al., 2012; Merchant and Honing, 2014), respond more accurately to temporally regular than jittered sequences (Zarco et al., 2009), based on their isochrony, rather than on beat-based perception (Merchant et al., 2015a; Merchant and Bartolo, 2017).

Based on these and related neurobiological observations (e.g., Zarco et al., 2009; Merchant et al., 2011) the *GAE* hypothesis was proposed (Merchant and Honing, 2014), arguing that the integration of sensorimotor information throughout the mCBGT circuit and other brain areas during the perception or execution of single intervals is similar in human and non-human primates, but different in the processing of multiple intervals. While the mCBGT circuit was shown to be also involved in beat-based mechanisms in imaging studies (e.g., Teki et al., 2011), direct projections from the medial premotor cortex (MPC) to the primary auditory cortex (A1) via the inferior parietal lobe (IPL) that is involved in sensory and cognitive functions such as attention and spatial sense (see **Figure 1**), may be the underpinning of beat-based timing as found in humans, and possibly apes. The GAE hypothesis suggests beat-based timing to be more developed in humans as opposed to apes and monkeys, and that it evolved through a gradual chain of anatomical and functional changes to the interval-based mechanism to generate an additional beat-based mechanism, instead of claiming a categorical jump from single-interval to multiple-interval abilities (i.e., rhythmic entrainment; Patel, 2006; Patel and Iversen, 2014). As such, the GAE hypothesis suggests that beat perception and entrainment have emerged gradually in primate order. This observation is in line with Rauschecker and Scott (2009) earlier suggestion that "the privileged access of the humans' auditory system to the sequential and temporal machinery of the mCBGT circuit emerged gradually in the course of evolution from precursors of the great ape lineage." Some recent behavioral studies support such a gradual interpretation (Hattori et al., 2015; Large and Gray, 2015) suggesting at least some beat-based timing capabilities in apes that are absent in rhesus monkeys. Finally, the GAE hypothesis is in line with Patel and Iversen (2014), that argue for a causal link between auditory and motor planning regions needed for human beat perception. However, the GAE hypothesis differs from the latter proposal in that it (a) does not claim the neural circuit that is engaged in beat-based timing to be deeply linked to vocal learning, perception,

**TABLE 1 |** Homologies between rhesus monkey, chimpanzee, and human cortical auditory evoked potentials (ERPs).

| | Human scalp Picton et al. | Ape scalp Ueno et al. | Monkey scalp Itoh et al. | Monkey scalp Gil-da-Costa et al. | Monkey scalp Honing et al. | Monkey cranial Teichert | Monkey epidural Javitt et al. |
|---|---|---|---|---|---|---|---|
| P1 | 50–60 | | 25–30 [mP1] | – | – | 45–65 [P55] | 5–40 |
| N1 | 75–100 | | 45–65 [mN1] | – | – | 70–105 [N85] | 40–120 |
| MMN | 100–200 | 125–180 | – | 48–120 | 60–110 | – | – |
| P3a | 200–250 | | – | 100–250 | – | – | – |

*Time range in ms; alternative naming in square brackets (Adapted from Picton et al., 1974; Javitt et al., 2000; Ueno et al., 2008; Honing et al., 2012; Gil-da-Costa et al., 2013; Itoh et al., 2015; Teichert, 2016).*

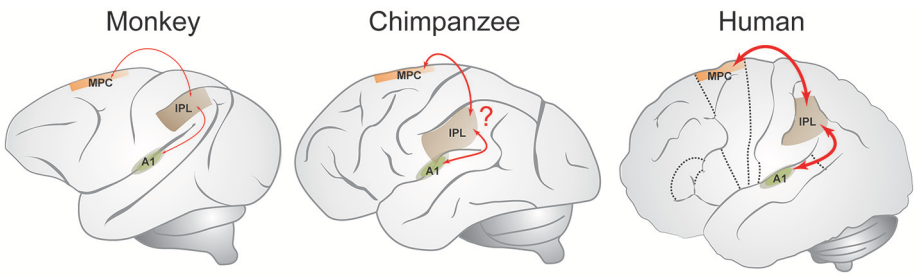



**FIGURE 1 |** The gradual audiomotor evolution (GAE) hypothesis. The GAE hypothesis suggests connections between medial premotor cortex (MPC), inferior parietal lobe (IPL), and primary auditory area (A1) to be stronger in humans as compared to other primates (marked with red lines), suggesting beat-based timing to have gradually evolved. Line thickness indicates the hypothesized connection strength (Adapted from Mendoza and Merchant, 2014 and Merchant and Honing, 2014).

and production, even if some explicit overlap between these neural circuits exists, and (b) that it gradually evolved in primates, instead of being solely present in humans as the only primate capable of vocal learning (Honing and Merchant, 2014).

However, in the current study we will not make claims about the underlying neural mechanisms, nor will we present a systematic comparative study (this is a topic of ongoing research). Instead, we will focus on whether rhesus monkeys are able to sense isochrony and/or the beat in a rhythmic stimulus.

## CURRENT STUDY

In the current study we adapted an auditory oddball paradigm that was previously used in humans (Bouwer et al., 2016) and that allows for testing and dissociating the contributions of beat perception and isochrony to the temporal predictability of the stimulus.

We presented two rhesus monkeys (*M. mulatta*) a rhythmic sequence that was made up of a pattern of loud and soft percussive sounds such that the acoustic stimulus could *induce* a simple binary metrical structure (duple meter), with an accented beat on every other metrical position (see combinations of S1 and S2 in **Figure 2**). This rhythmic sequence was presented in two conditions: an *isochronous condition*, in which the sounds were presented in an isochronous fashion, using an inter-onset interval (IOI) of 225 ms, allowing a beat to be induced (i.e., one metrical level of a duple meter). And a *jittered condition*, in which the IOIs were randomly selected between 150 and 300 ms, as such disabling the perception of a beat. Furthermore, we used intensity decrements to be able to compare the ERP response to deviants in both the isochronous and jittered conditions (since omissions, as used in Honing et al., 2012, wouldn't be recognized as deviants in the jittered condition). By introducing unexpected intensity decrements (i.e., deviants) on both on the beat and offbeat positions, in both the isochronous and jittered conditions, we could probe the effect of metrical position (beat vs. offbeat), as well as the effect of isochrony (isochronous vs. jittered) on the amplitude of the MMN and the P3a. Additionally, we examined the effects of metrical position and isochrony on P1 and N1 responses to standard sounds.

This design allows for testing several hypotheses. First of all, we expected an MMN and P3a for all deviants in both the isochronous and the jittered conditions, irrespective of isochrony or metrical position. This to make sure that the auditory system of monkeys is sensitive to unexpected amplitude decrements (deviants) in a rhythmic stream. For this we predicted an effect of Type (standard vs. deviant).

Second, we did not expect to find evidence for beat perception, in line with earlier findings (Honing et al., 2012; Merchant and Honing, 2014). As such, we predicted no interaction between Position (Beat vs. Offbeat) and Isochrony (Isochronous vs. Jittered). To show beat perception, the difference between the MMN responses to deviants on the beat and offbeat in the isochronous condition should be more pronounced than the difference between the MMN responses to deviants on beat and offbeat positions in the jittered condition, in which beat perception is disabled. Without such an interaction between metrical position and isochrony, the differences between the MMN responses to different metrical positions should be

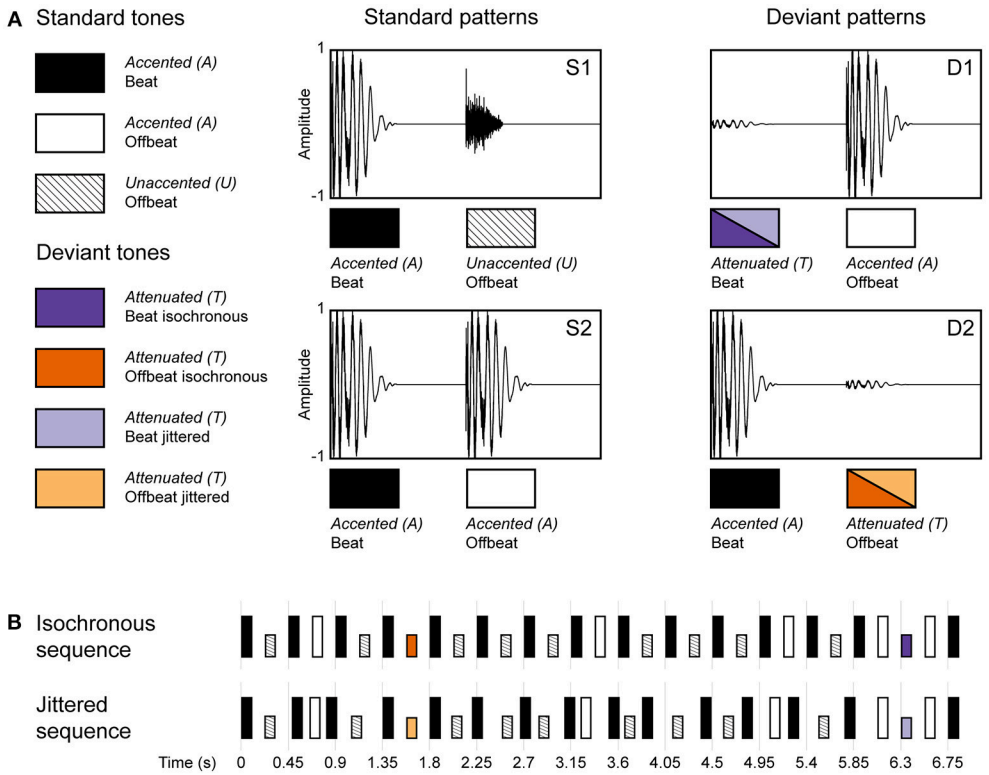



**FIGURE 2 |** Schematic diagram of the rhythmic stimulus patterns used in the experiment. **(A)** The two standard (S1 and S2) and two deviant patterns (D1 and D2), made up of three different sounds (A, accented; U, unaccented; and T, attenuated). An accented sound could occur either on the beat or offbeat, an unaccented sound was restricted to the offbeat position. An attenuated sound was used as a deviant in two positions (beat and offbeat) and in two conditions (isochronous and jittered). **(B)** Standard and deviant sound patterns were concatenated into a single rhythmic stream (their generation is visualized as a transition network in **Figure 3**). Sequences in the isochronous condition had an inter-onset interval (IOI) of 225 ms, in the jittered condition these were randomly chosen from the range 150–300 ms using a uniform distribution. Deviants were always preceded and followed by an accented sound, with a fixed IOI of 225 ms in both conditions (Adapted from Bouwer et al., 2016).

interpreted as a result of sequential learning instead of beat perception (see Bouwer et al., 2016).

Third, we did expect the MMN and the P3a to be affected by the isochrony of the stimulus, with both having a higher amplitude in the isochronous condition (isochronous rhythm) as compared to the jittered condition (random rhythm). When monkeys are sensitive to the temporal regularity of the isochronous stimulus (Merchant et al., 2015b; Gámez et al., 2018) this should help in predicting the next event (i.e., increasing its temporal predictability) and enhancing its processing (Schwartze et al., 2011; Bouwer et al., 2014). Hence, it can be expected that the amplitude of the MMN and P3a in response to deviants in the isochronous condition is larger than in the jittered condition. As such, we predicted an interaction between Isochrony (Isochronous vs. Jittered) and Type (Standard vs. Deviant). In addition, we expected an inverse relationship between the amplitude of the P1 and N1 and temporal predictability. When the amplitude of the N1 and P1 is attenuated by the isochrony of the stimulus this can be used as additional evidence for isochrony perception (as was shown for humans in Schwartze et al., 2013).

# METHODS

## Ethics Statement

All the animal care, housing, experimental procedures were approved by the National University of Mexico Institutional Animal Care and Use Committee and conformed to the principles outlined in the Guide for Care and Use of Laboratory Animals (NIH, publication number 85–23, revised 1985). Both monkeys were monitored daily by the researchers and the animal care staff, and every second day from the veterinarian, to check the conditions of health and welfare. To ameliorate their condition of life we routinely introduced in the home cage (1.3 $m^3$) environment toys (often containing items of food that they liked) to promote their exploratory behavior. The researcher that tested the animals spent half an hour interacting with the monkeys directly, giving for example new objects to manipulate. We think that this interaction with humans, in addition to the interaction that was part of the task performed, can help to reduce potential stress related to the experiment. Food and water where given ad libitum.

## Participants

Two rhesus monkeys participated in the ERP measurements. Monkey A is a 11 year old male, Monkey B a 9 year old male. Both monkeys have normal hearing. They were awake (i.e., not sedated) during the measurements, sitting in a quiet room [3 (l) × 2 (d) × 2.5 (h) m] with dimmed lighting and two loudspeakers in front of them. The ERP measurements were performed after a morning session of unrelated behavioral experiments. The animals were seated comfortably in a monkey chair where they could freely move their head, hands and feet. No head fixation was used and the EEG electrodes were attached to the monkey's scalp using tape. To ease the fixation of the electrodes, the monkey's hair on the scalp and reference ear was shaved.

## Stimuli

The stimuli were identical to those used in Bouwer et al. (2016). Rhythmic sequences were composed of two sounds that differed in timbre, intensity and duration to induce a simple binary metrical structure (duple meter) with acoustic accents. The sounds were made with QuickTime's (Apple, Inc.) drum timbres. The first sound consisted of a simultaneously sounding bass drum and hi-hat, and will be referred to as *accented* (or A for short). The second sound was a hi-hat, which was 16.6 dB softer than the accented sound and lasted 70 ms instead of 110 ms. This sound will be referred to as *unaccented* (or U for short). The deviant sound was created by attenuating the accented sound by 25 dB (using Praat software; www.praat.org), leaving timbre and duration intact. This sound will be referred to as *attenuated* (or T for short; see **Figure 2**).

The accented, unaccented and attenuated sounds (A, U, and T) were combined into a rhythmic stream in which 60% of the time an accented sound was followed by an unaccented sound (see S1 in **Figure 2A**), and 30% of the time an accented sound was followed by another accented sound (see S2 in **Figure 2A**), as such inducing a duple meter, with always an accented sound on the beat. In the remaining 10% of the time a deviant was inserted (the "oddball"). This was either, randomly chosen, an attenuated sound followed by an accented sound (5%; see D1 in **Figure 2A**) or an accented sound followed by attenuated sound (5%; see D2 in **Figure 2A**). An example of an isochronous and a jittered rhythmic stream are given in the Supplementary Material.

The resulting sequence was used in both the isochronous and jittered conditions (see **Figure 2B**). In the isochronous condition, all sounds were presented with a constant inter-onset interval of 225 ms. In this condition, the probabilistic pattern of alternating accented and unaccented sounds was expected to induce a beat (or duple meter) with an inter-beat interval of 450 ms, within the optimal range for beat perception in humans (London, 2002). Sounds in uneven positions of the sequence (including deviant D1) can be considered on the beat, while all sounds in even positions (including deviant D2) are offbeat. In the jittered condition, the inter-onset intervals were randomly distributed between 150 and 300 ms with an average of 225 ms (uniform distribution), using the same sequence as in the isochronous condition, making it impossible to induce a regular beat (London, 2002; Honing, 2012). However, the inter-onset interval just before and after a deviant tone was kept constant at 225 ms. This was done to make the acoustic and temporal context in which a deviant occurs, identical between both conditions.

In addition, four additional constraints were applied to the construction of the sound sequences. To optimize the possibility of inducing a beat in the isochronous condition, S2 (containing two consecutive accented sounds) was never presented more than once in a row, and only a maximum of four consecutive S1 patterns was allowed. Furthermore, a deviant on the beat (D1) was always preceded by an accented sound offbeat (S2), ensuring the acoustical context to be identical for all deviants. Finally, at least five standard patterns occurred between two deviant patterns. For schematic examples of both the isochronous and jittered sequences, see **Figure 2B**. Note that the D1 and D2 are referred to as on the beat or offbeat in both conditions for comparison, while this can only be perceived as such in the isochronous condition.

The statistical properties of the sequences used are visualized in **Figure 3** as a transition network, with the three basic sounds A, U, and T as nodes. Note that this is a simplification in the sense that it does not include the four constraints mentioned above.

## Procedure

Stimuli were presented in blocks consisting of 3 isochronous and 3 jittered sequences (1,300 pattern, i.e., 2,600 sound events, per sequence), in randomized order. One block (of 6 sequences) lasted 9 min and 45 s, separated by a silent interval of about 15 s. Isochronous and jittered blocks were presented in semi-random order, with a maximum of two blocks from the same condition following each other. This resulted in a total session length of 60 min, one session per day.

Sound stimuli were presented through 2 loudspeakers placed 1.1 meters away from the subject (and 1 meter apart from each other). The sound intensity measured at the subject position was 80 dB SPL. The monkeys participated in one recording session per day, to a total of 14 sessions for Monkey A and 12 sessions for Monkey B. All measurements were completed in about 5 weeks per monkey.

Overall, the design of the study was identical to the unattended condition presented in Bouwer et al. (2016), except that in the human study participants watched a silent video with subtitles, whereas in the current study monkeys were not given any visual stimuli to focus on.

## EEG Recording and Analysis

The EEG was recorded from electrodes (Grass EEG electrodes; #FS-E5GH-60) attached to five scalp positions (Fz, Cz, Pz, F3, F4) according to the 10–20 system (see **Figure 4**).

The electrodes were connected to a Tucker-Davis Technologies (TDT) headstage (#RA16LI) for low impedance electrodes. This headstage was connected to a TDT RA16PA preamplifier, which in turn was connected to a TDT RZ2 processor. RZ2 was programmed to acquire the EEG signals with a sampling rate of 498.25 Hz and the bandpass filters were set at 0.01–100 Hz.

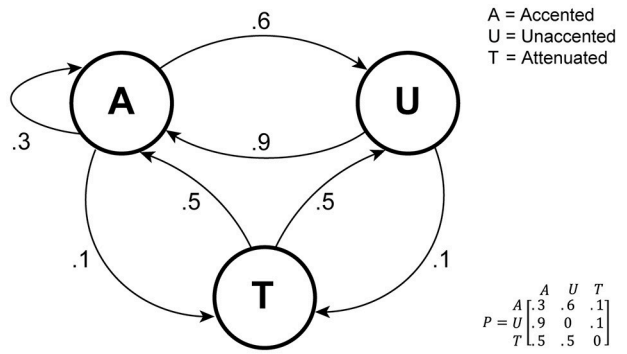



**FIGURE 3 |** Rhythmic sequence generation visualized as a transition network. As noted in the text, this representation is a simplification in that it does not include certain constraints on which patterns can follow each other.

All electrodes were attached using Ten20 Conductive EEG Paste and medical tape, and were referenced to the right ear (fleshy part of the pinna). In the offline analysis, a 0.1–30 Hz band-pass FIR filter (Kaiser-window) was applied.

## MMN and P3a

For the analyses of MMN and P3a, epochs of −150 to 300 ms were extracted for the four deviant patterns ($D1_i$, $D1_j$, $D2_i$, $D2_j$; with subscripts referring to isochronous and jittered conditions). Epochs of the same length were extracted for accented sounds from the standards in the isochronous condition, both on the beat (from S1, but only if preceded by S2) and offbeat (from S2). Thus, the acoustic context preceding all sounds that were used in the analysis of the MMN and P3a was identical (i.e., they were preceded by an accented sound at −225 ms). Epochs were baseline corrected using the average voltage of the 150 ms prior to the onset of the tone and averaged to obtain ERPs for each condition and monkey. We obtained difference waves by subtracting the ERP responses to the accented sounds from the standard patterns from the ERP responses to the deviant tones at the same position (beat or offbeat). Epochs that exceeded ±300 μV amplitude were excluded from the statistical analysis. The number of epochs accepted for analysis are given in **Tables 2**, **3**.

We defined the amplitude of the MMN as the average amplitude from a 30 ms window (relatively small to avoid overlap with the N1) centered around the average peak latency across conditions on Cz (electrode that was shown to be maximally indicative of MMN in rhesus monkeys; Honing et al., 2012; Gil-da-Costa et al., 2013). The MMN peaked at Cz on average at 72 ms for Monkey A and at 110 ms for Monkey B. See the caption of **Table 2** for the time windows used in the statistical analyses.

We defined the amplitude of the P3a as the average amplitude from a 50 ms window centered around the average peak latency across conditions on Cz (electrode that was shown to be maximally indicative of P3a in rhesus monkeys; Gil-da-Costa et al., 2013). The P3a peaked at Cz on average at 151 ms for Monkey A and at 203 ms for Monkey B. See the caption of **Table 3** for the time windows used in the statistical analyses.

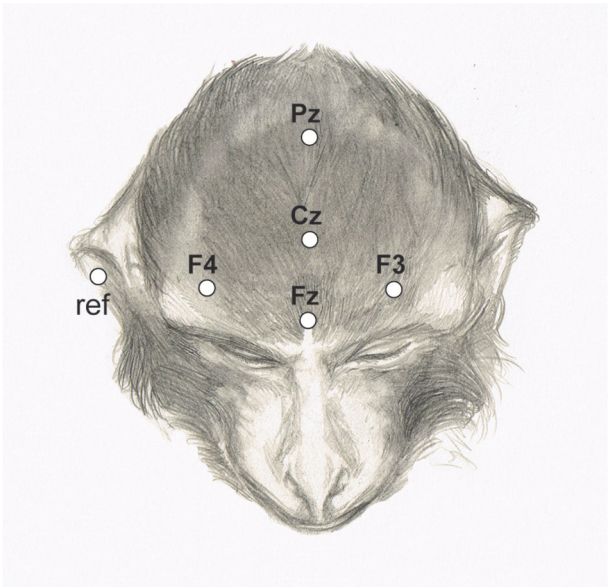



**FIGURE 4 |** Electrode positions marked on the head of a rhesus monkey. (Adapted from Honing et al., 2012).

**TABLE 2 |** MMN.

| **Monkey A** | | | | |
|---|---|---|---|---|
| | **Beat** | | **Offbeat** | |
| | **D (*n* = 2,695)** | **S (*n* = 13,567)** | **D (*n* = 2,708)** | **S (*n* = 11,683)** |
| Isochronous | −0.37 (0.27) | 1.60 (0.12) | −0.001 (0.27) | 1.34 (0.13) |
| | (n = 2,719) | (n = 13,567) | (n = 2,724) | (n = 11,683) |
| Jittered | 0.30 (0.27) | 1.60 (0.12) | 0.50 (0.28) | 1.34 (0.13) |
| **Monkey B** | | | | |
| | **Beat** | | **Offbeat** | |
| | **D (*n* = 2,125)** | **S (*n* = 10,544)** | **D (*n* = 2,104)** | **S (*n* = 9,176)** |
| Isochronous | −0.58(0.31) | 1.10 (0.14) | −0.38 (0.31) | 0.95 (0.15) |
| | (n = 2,250) | (n = 10,544) | (n = 2,273) | (n = 9,176) |
| Jittered | 0.39 (0.31) | 1.10 (0.14) | 0.51 (0.29) | 0.95 (0.15) |

*Mean amplitudes of standard and deviant waves at Cz in the isochronous and jittered conditions for Monkey A and Monkey B. Mean amplitudes (μV) are indicated with SE values in parentheses. S, values for standard stimuli; D, values for deviant stimuli; n, number of epochs. The time window used in the statistical analyses is 57–87 ms for Monkey A and 95–125 ms for Monkey B.*

## P1 and N1

For the analysis of P1 and N1 epochs of −100 to 300 ms were extracted for accented standards on the beat and offbeat, including only those standards that were preceded by an accented sound. A filter of 5–75 Hz was applied to eliminate slow drift (as is commonly used in studying these components in humans; see Schwartze et al., 2013; Bouwer and Honing, 2015). Hence, baseline correction was not needed. Epochs that exceeded ±300 μV amplitude were excluded from the statistical analysis.

**TABLE 3 |** P3a.

**Monkey A**

| | Beat | | Offbeat | |
|---|---|---|---|---|
| | **D (*n* = 2,695)** | **S (*n* = 13,567)** | **D (*n* = 2,708)** | **S (*n* = 11,683)** |
| Isochronous | 0.78 (0.21) | −1.51 (0.10) | 1.23 (0.23) | −1.72 (0.11) |
| | (n = 2,719) | (n = 13,567) | (n = 2,724) | (n = 11,683) |
| Jittered | −0.06 (0.22) | −1.51 (0.10) | 0.03 (0.22) | −1.72 (0.11) |

**Monkey B**

| | Beat | | Offbeat | |
|---|---|---|---|---|
| | **D (*n* = 2,125)** | **S (*n* = 10,544)** | **D (*n* = 2,104)** | **S (*n* = 9,176)** |
| Isochronous | 0.21 (0.27) | −0.95 (0.13) | 0.32 (0.28) | −0.96 (0.13) |
| | (n = 2,250) | (n = 10,544) | (n = 2,273) | (n = 9,176) |
| Jittered | 0.58 (0.27) | −0.95 (0.13) | 0.35 (0.24) | −0.96 (0.13) |

*Mean amplitudes of P3a at Cz in the isochronous and jittered conditions. Mean amplitudes (μV) are indicated with SE values in parentheses. Number of epochs: n between brackets. The time window used in the statistical analyses is 126–176 ms for Monkey A and 178–228 ms for Monkey B.*

**TABLE 4 |** P1 and N1.

**Monkey A**

| | Beat | | Offbeat | |
|---|---|---|---|---|
| | **P1 (*n* = 13,598)** | **N1 (*n* = 13,598)** | **P1 (*n* = 11,707)** | **N1 (*n* = 11,707)** |
| Isochronous | 0.83 (0.12) | −2.24 (0.12) | 0.64 (0.13) | −2.37 (0.13) |
| | (n = 13,494) | (n = 13,494) | (n = 11,774) | (n = 11,773) |
| Jittered | 1.28 (0.13) | −3.08 (0.13) | 1.65 (0.14) | −2.96 (0.14) |

**Monkey B**

| | Beat | | Offbeat | |
|---|---|---|---|---|
| | **P1 (*n* = 10,734)** | **N1 (*n* = 10,734)** | **P1 (*n* = 9,307)** | **N1 (*n* = 9,307)** |
| Isochronous | 2.85 (0.10) | −1.72 (0.10) | 2.96 (0.11) | −1.62 (0.11) |
| | (n = 11,224) | (n = 11,224) | (n = 9,750) | (n = 9,750) |
| Jittered | 2.27 (0.10) | −1.74 (0.10) | 2.43 (0.11) | −1.59 (0.11) |

*Mean amplitudes of P1 and N1 at Cz in the isochronous and jittered conditions. Mean amplitudes (μV) are indicated with SE values in parentheses. Number of epochs: n between brackets. The time window used in the statistical analyses for Monkey A is 20–40 ms for the P1 and 40–60 ms for the N1. For Monkey A the time window used is 18–38 ms for the P1 and 40–60 ms for the N1.*

The peak amplitude of both the P1 and N1 was defined on the average of the waveforms collapsed over conditions, using 20 ms window centered around the average peaks at Cz (electrode that was also shown in other studies to be maximally indicative of P1 and N1 in monkeys; see Itoh et al., 2015). A 20 ms window was chosen around the average P1 and N1 peaks to avoid overlap between the P1 and N1 windows. **Table 4** shows the average amplitudes for all four conditions, the number of epochs accepted for analysis, and the time windows used for statistical analyses.

## Statistical Analysis

The amplitudes extracted from the difference waves were entered into ANOVAs with factors Position (Beat vs. Offbeat), Isochrony (Isochronous vs. Jittered), and Type (Standard vs. Deviant) for the MMN and P3a analyses, and the factors Position (Beat vs. Offbeat), Isochrony (Isochronous vs. Jittered) for the P1 and N1 analyses. Partial eta squared ($\eta_p^2$) was used as a measure of effect size. All statistical analyses were conducted in SPSS (Version 22).

# RESULTS

## MMN and P3a

**Figures 5**, **6** show that, for both monkeys, the ERPs elicited by the standard (dotted lines) and the deviant (solid lines) are different, with peaks in the interval 60–110 ms for the MMN and 100–250 ms for the P3a, consistent with earlier studies (see **Table 1**).

For both monkeys the ANOVA with factors Type (Standard vs. Deviant), Isochrony (Isochronous vs. Jittered) and Position (Beat vs. Offbeat) revealed a significant main effect of Type {Monkey A: $[F_{(1,\,61345)} = 83.790,\ p < 0.0005,\ \eta_p^2 = 0.001]$; Monkey B: $[F_{(1,\,48191)} = 38.906,\ p < 0.0005,\ \eta_p^2 < 0.001]$}, showing that the evoked brain response to deviants was significantly negative as compared to the standard in all conditions. In addition, there was an interaction of Isochrony and Type {Monkey A: $[F_{(1,\,61345)} = 3.896,\ p < 0.048,\ \eta_p^2 = 0.0005]$; Monkey B: $[F_{(1,\,48191)} = 7.819,\ p < 0.005,\ \eta_p^2 < 0.0005]$} showing an effect of isochrony on the size of the MMN, being larger in the isochronous condition. This suggests a sensitivity to the isochrony of the stimulus. However, no effects of Position or an interaction between Position and Isochrony were found. Hence, there is no support for beat perception. See **Table 2** for all MMN measurements and **Figure 8** for a summary.

With regard to the P3a there is a significant main effect of Type {Monkey A: $[F_{(1,\,61345)} = 292.069,\ p < 0.0005,\ \eta_p^2 = 0.005]$; Monkey B: $[F_{(1,\,48191)} = 76.793,\ p < 0.0005,\ \eta_p^2 < 0.002]$} and an interaction of Isochrony and Type {Monkey A: $[F_{(1,\,61345)} = 17.032,\ p < 0.0005,\ \eta_p^2 = 0.0005]$; Monkey B: n.s.}. For Monkey A there was also a significant interaction between Position and Type {Monkey A: $[F_{(1,\,61345)} = 3.884,\ p < 0.049,\ \eta_p^2 < 0.0005]$}. Note that this is in a direction opposite to what was found in humans (Bouwer et al., 2016), which makes this result difficult to interpret. See **Table 3** for all P3a measurements and **Figure 8** for a summary.

## P1 and N1

**Figure 7** shows the P1 and N1 responses to the standard in the isochronous condition (solid lines) and jittered condition (dotted lines) for both monkeys. **Table 4** shows the average amplitudes for all four conditions and the time windows used for statistical analyses.

For both monkeys the ANOVA with factors Isochrony (Isochronous vs. Jittered) and Position (Beat vs. Offbeat) revealed

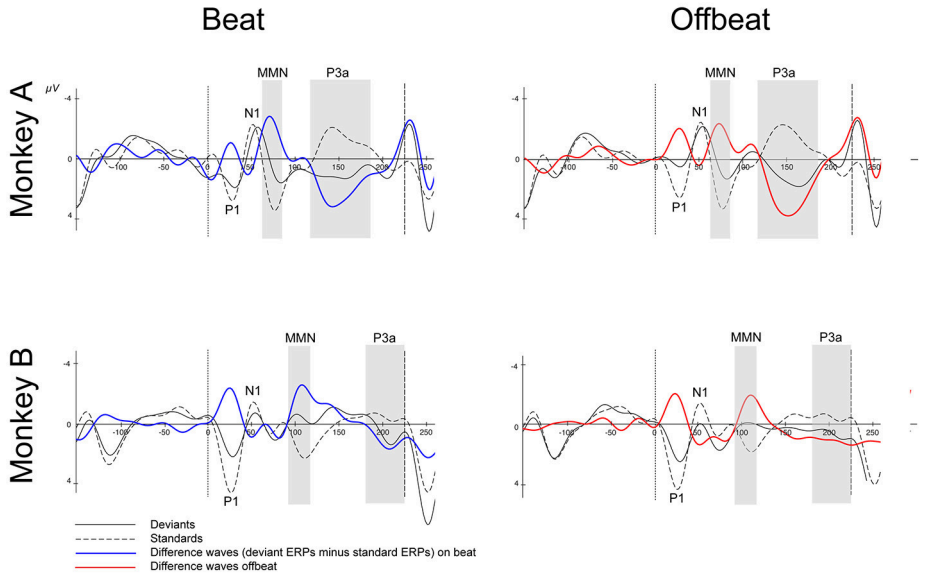



**FIGURE 5 |** Event-related potentials at Cz for the isochronous condition. Deviants are marked with solid black lines, standards with dotted black lines. Difference waves (deviant ERPs minus standard ERPs) on the beat are represented with solid blue lines, offbeat with solid red lines, for Monkey A and Monkey B. The gray-shaded areas indicate the time windows used in the statistical analyses of the MMN and P3a. Dotted vertical lines indicate the start of the sound event, dashed vertical lines indicate the onset of the next sound event (at 225 ms). See **Table 2** for details on time ranges used.

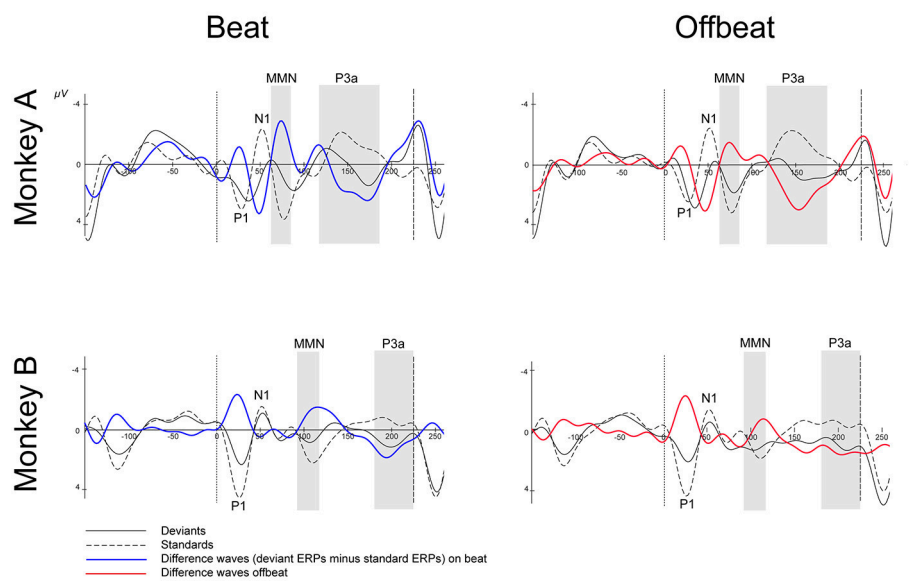



**FIGURE 6 |** Event-related potentials at Cz for the jittered condition. Deviants are marked with solid black lines, standards with dotted black lines. Difference waves (deviant ERPs minus standard ERPs) on the beat are represented with solid blue lines, offbeat with solid red lines, for Monkey A and Monkey B. The gray-shaded areas indicate the time windows used in the statistical analyses of the MMN and P3a. Dotted vertical lines indicate the start of the sound event, dashed vertical lines indicate the onset of the next sound event (at 225 ms). See **Table 3** for details on time ranges used.

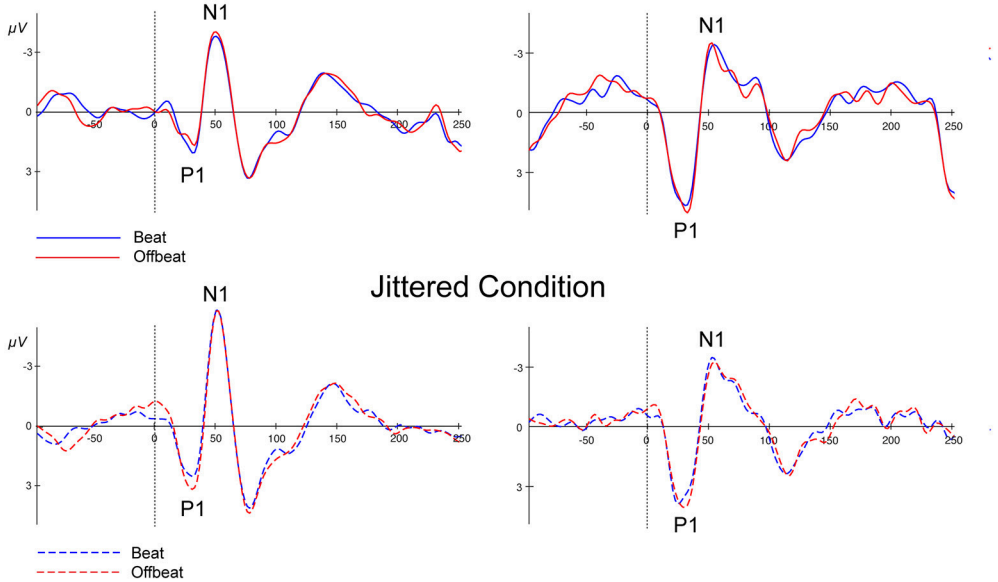



**FIGURE 7 |** Event-related potentials at Cz for all conditions. Standards in the isochronous condition **(Top)** and jittered condition **(Bottom)** for Monkey A and Monkey B. Standards on the beat are marked in blue and standards offbeat are marked in red.

a significant main effect of Isochrony for both the P1 {Monkey A: $[F_{(1,\ 50572)} = 31.306,\ p < 0.0005,\ \eta_p^2 = 0.001]$; Monkey B: $[F_{(1,\ 41014)} = 27.912,\ p < 0.0005,\ \eta_p^2 = 0.001]$} and the N1 {Monkey A: $[F_{(1,\ 50572)} = 30.822,\ p < 0.0005,\ \eta_p^2 = 0.001]$; Monkey B: n.s.} showing that the P1 is significantly different in the jittered as compared to the isochronous condition for both monkeys, and the N1 only for Monkey A. However, in the case of the P1 the direction of the effect is different for both monkeys. In Monkey A both the P1 and the N1 are larger for the jittered as compared to the isochronous sounds. This is in line with the idea that unpredictable stimuli receive more processing, hence resulting in larger amplitude responses (Schwartze et al., 2011, 2013). However, in Monkey B the effect of Isochrony on the P1 response is in the opposite direction, with larger responses to the sounds in the isochronous as opposed to in the jittered sequences. As such, this weakens the interpretation that the P1 is indicative of predictability. Additionally, we found an interaction between Isochrony and Position in Monkey A $\{[F_{(1,\ 50572)} = 4.585, p < 0.032, \eta_p^2 < 0.0005]\}$. However, the latter is in the opposite direction as compared to humans (Baldeweg, 2007; Costa-Faidella et al., 2011) and hence hard to interpret. All other effects were not significant.

## DISCUSSION

In the current study we examined the role of beat perception and isochrony perception in two rhesus monkeys using the same stimuli in an oddball paradigm that was previously used in humans (Bouwer et al., 2016). Similar to humans, we found MMNs to all deviants as opposed to standards in both the isochronous and jittered conditions. From this we conclude that the monkey brain is sensitive to unexpected amplitude decrements (deviants). This is in line with earlier studies that used either omissions (Honing et al., 2012) or intensity deviants (Gil-da-Costa et al., 2013).

However, and contrary to what was found in humans (Bouwer et al., 2016), there are no significant differences in the MMN between deviants in beat positions as opposed to deviants in an offbeat position, neither in the isochronous condition nor the jittered condition (i.e., no effect of metrical position, nor an interaction with regularity; see **Figure 8**). So while in humans the beat appears to modulate the amplitude of the MMN, in monkeys there was not such an effect. This suggests the absence of beat perception in monkeys, in line with an earlier study (Honing et al., 2012).

With regard to the P3a, Monkey A did show an interaction between Type and Isochrony, but this was small and in the opposite direction as compared to humans (see **Figure 8**). This makes the latter result difficult to interpret. The effect might be a result of overlapping components (e.g., an interference with the strong response to the next onset; marked with dashed lines in **Figures 5**, **6**), as such affecting the overall amplitude of the ERP signal. It could also be a result of attentional fluctuations that are modulating the P3a, as was shown in human studies (cf. Bouwer et al., 2016). Future work should manipulate attention to make sure whether and how attention influences rhythm perception.

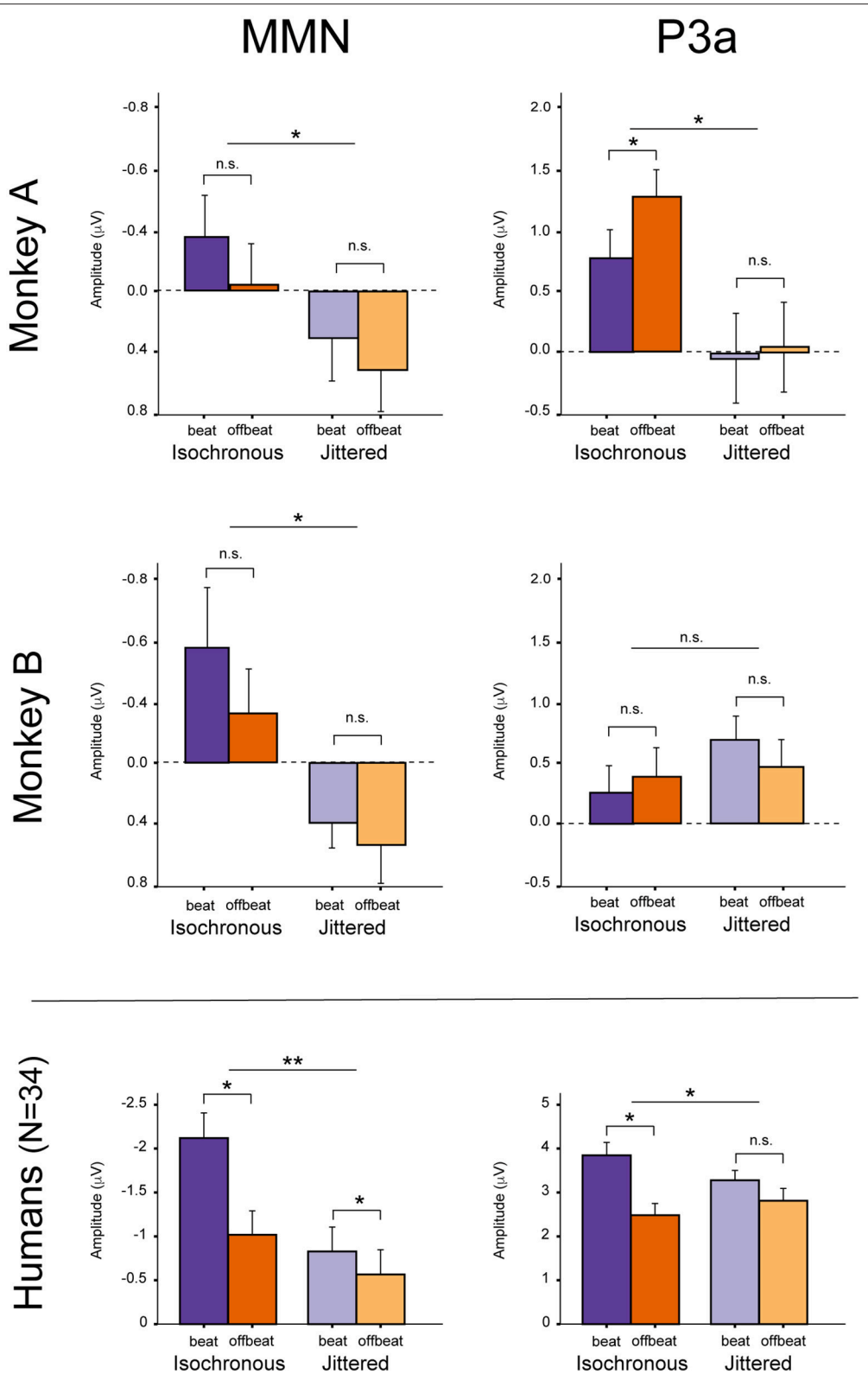



**FIGURE 8 |** Summary of the results. MMN and P3a in Monkey A **(Top)** and Monkey B **(Middle)**. For comparison, the **(Bottom)** shows the mean results for humans ($N = 34$), unattended condition (see Table 1 in Bouwer et al., 2016). Note that the y-axes are different for the monkey and human data. Error bars represent one standard error of the mean. *Significant interaction.

**TABLE 5 |** Summary of the results.

| | Deviants | | Standards | |
|---|---|---|---|---|
| | **MMN** | **P3a** | **P1** | **N1** |
| Monkey A | * | * | * | * |
| Monkey B | * | n.s. | – | n.s. |

*The effect of isochrony on MMN and P3a (enhancement), and temporal predictability on P1 and N1 (attenuation). *,significant; n.s., non-significant; –, significant in opposite direction.*

By contrast, both monkeys appear to be sensitive to the isochrony of the stimulus, as is supported by an overall larger MMN response to deviants in the isochronous as opposed to the jittered condition (see **Figure 8**). Note however, that we cannot make the claim that monkeys detect isochrony (hence the use of the term "sense"). It could well be that the monkey brain is able to sense isochrony, but that the monkey is not aware of this, as was recently shown in two beat-deaf humans (Mathias et al., 2016). The latter study suggested the P3a to be associated with conscious perception or "awareness," comparable to what is observed in tone deaf humans (Peretz, 2016).

With regard to the N1 and P1 the results are mixed and not consistent between Monkey A and Monkey B. While the attenuation of the N1 of Monkey A is in support for isochrony perception, the P1 is in opposite directions for the two monkeys. This difference could again be caused by attention, which was not controlled for in this study. As such we have to be cautious in our interpretation of the P1 and N1 in the standards, and will base our conclusions mainly on the MMN in response to the deviants (see summary in **Table 5** and main results in **Figure 8**).

In short, the ERPs of both monkeys appear to be influenced by isochrony, but not by the induced beat. Where humans show a clear interaction between metric position and isochrony (Bouwer et al., 2016), and as such show evidence for beat perception, the current results provide no evidence for beat perception in rhesus monkeys. Hence, the hypothesis that beat-based timing is common to all animals (Darwin, 1871; Wilson and Cook, 2016) is not supported by these results.

While the underlying neural mechanism (be it interval-based timing or otherwise; Teki et al., 2011, 2016) is as yet unclear, we take this result as further support for the GAE hypothesis (Merchant and Honing, 2014). The GAE hypothesis suggests beat-based timing to be a result of bidirectional bottom-up and top-down interactions between the auditory and motor areas of the brain, including the mCBGT circuit and parietal areas such as the IPL (see **Figure 1**), connections that are quite developed and efficient in humans and that emerged gradually in the course of evolution from precursors of the great ape lineage (Rauschecker and Scott, 2009; Honing and Merchant, 2014; Merchant and Honing, 2014; Morillon et al., 2014; Merchant and Yarrow, 2016).

Parts of the audiomotor circuit, including areas such as the putamen in the basal ganglia or the parietal cortex, also process information from the dorsal stream of visual processing (Kimura, 1992; Merchant et al., 2004). Therefore, it could well be that the areas involved in the strong visuomotor coupling in monkeys partially overlap with the beat-entrainment audiomotor system, in line with the predictive and entrainment abilities of monkeys with visual stimuli (Takeya et al., 2017; Gámez et al., 2018). Thus, where humans show a preference for auditory metronomes, monkeys have a clear preference for visual stimuli (Gámez et al., 2018). Applying the current paradigm to the visual modality (e.g., with different intensity flashes rather than different intensity tones) might be able to show beat-entrainment for visual stimuli.

Gámez et al. (2018) also provides evidence that monkeys can predictively entrain to an isochronous metronome, even when it accelerates or decelerates. These findings suggest that the beat-based mechanisms of macaques might not be as restricted as previously thought (Patel, 2014). Thus, it is crucial that future experiments focus on finding the limits in beat perception and entrainment capabilities of monkeys, using gradually more complex levels of metrical periodicity in their stimuli (Schwartze and Kotz, 2015; Bouwer et al., 2016).

Finally, and as predicted by the GAE hypothesis, we expect beat-based timing in the auditory modality to be present in some rudimentary form in apes (not monkeys). Two recent behavioral studies support this interpretation (Hattori et al., 2015; Large and Gray, 2015) for a chimpanzee (*P. troglodytes*) and a bonobo (*Pan paniscus*). Applying the same oddball paradigm to a chimpanzee (cf. Ueno et al., 2008) might find further support for the GAE hypothesis (Merchant and Honing, 2014). In addition, the paradigm could be extended to other species, for example, pigeons, rodents, cats and dogs (e.g., Nelken and Ulanovsky, 2007; Howell et al., 2012; Schall et al., 2015; Harms et al., 2016), where the MMN can be measured to probe the potentially shared capabilities of isochrony perception and/or beat perception.

While musicality is likely made up of many components, it appears to be good strategy to start with a focus on core aspects, like beat perception (cf. Honing, 2018b). The core aspects of musicality are well suited for comparative studies, both cross-cultural and cross-species, and the nature and extent of their presence in non-human animals have attracted considerable debate in the recent literature. These recent discussions, combined with the availability of suitable experimental techniques for tracking these phenomena in human and non-human animals, make this a timely and feasible enterprise. Of course, we need to remain cautious about making claims on music-specific modes of processing until more general accounts have been ruled out. It still has to be demonstrated that the constituent components of musicality, when identified, are indeed domain specific. In contrast, the argument that music is a human invention (Patel, 2018) depends on the demonstration that the components of musicality are not domain specific, but each cognitively linked to some non-musical mental ability. So while there might be quite some evidence that components of musicality overlap with non-musical cognitive features (Patel, 2018), this is in itself no evidence against musicality as an evolved biological trait or

set of traits. As in language, musicality could have evolved from existing elements that are brought together in unique ways, and that system may still have emerged as a biological product through evolutionary processes, such as natural or sexual selection. As such there is no need for musicality to be modular or show a modular structure. Alternatively, based on the converging evidence for music-specific responses along specific neural pathways, it could be that brain networks that support musicality are partly recycled for language, thus predicting more overlap than segregation of cognitive functions. In fact, this is one possible route to test the Darwin-inspired conjecture that musicality precedes music and language (Honing, 2018a).

## AUTHOR CONTRIBUTIONS

HH and FB conceived and designed the experiments. HM and LP performed the experiments. FB and HH analyzed the data. HH, FB, and HM wrote the paper.

## ACKNOWLEDGMENTS

Thanks to Germán Mendoza for assisting in running the monkey experiments. Yaneri Ayala, Yonathan Fishman, Yukiko Kikuchi, Sonja Kotz, Ani Patel, István Winkler and four reviewers are thanked for their constructive criticisms on an earlier version of the manuscript. The first author HH is supported by a Horizon grant (317-70-010) of the Netherlands Organization for Scientific Research (NWO). HM is supported by CONACYT: 236836, CONACYT: 196, and PAPIIT: IN201214-25.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2018.00475/full#supplementary-material

**Audio 1 |** Example of a rhythmic pattern used in the isochronous condition. [1_isochronous.wav].

**Audio 2 |** Example of a rhythmic pattern used in the jittered condition. [2_jittered.wav].

## REFERENCES

Baldeweg, T. (2007). ERP repetition effects and mismatch negativity generation: a predictive coding perspective. *J. Psychophysiol.* 21, 204–213. doi: 10.1027/0269-8803.21.34.204

Bouwer, F. L., and Honing, H. (2015). Temporal attending and prediction influence the perception of metrical rhythm: evidence from reaction times and ERPs. *Front. Psychol.* 6:1094. doi: 10.3389/fpsyg.2015.01094

Bouwer, F. L., van Zuijen, T. L., and Honing, H. (2014). Beat processing is pre-attentive for metrically simple rhythms with clear accents: an ERP study. *PLoS ONE* 9:e97467. doi: 10.1371/journal.pone.0097467

Bouwer, F. L., Werner, C. M., Knetemann, M., and Honing, H. (2016). Disentangling beat perception from statistical learning using ERPs: the role of attention and musical training. *Neuropsychologia* 85, 80–90. doi: 10.1016/j.neuropsychologia.2016.02.018

Cook, P., Rouse, A., Wilson, M., and Reichmuth, C. (2013). A California sea lion (*Zalophus californianus*) can keep the beat: motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *J. Comp. Psychol.* 127, 412–427. doi: 10.1037/a0032345

Costa-Faidella, J., Baldeweg, T., Grimm, S., and Escera, C. (2011). Interactions between "what" and "when" in the auditory system: temporal predictability enhances repetition suppression. *J. Neurosci.* 31, 18590–18597. doi: 10.1523/JNEUROSCI.2599-11.2011

Darwin, C. (1871). *The Descent of Man, And Selection in Relation to Sex*. London: John Murray.

Fishman, Y. I., and Steinschneider, M. (2012). Searching for the mismatch negativity in primary auditory cortex of the awake monkey: deviance detection or stimulus specific adaptation? *J. Neurosci.* 32, 15747–15758. doi: 10.1523/JNEUROSCI.2835-12.2012

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Front. Syst. Neurosci.* 7:68. doi: 10.3389/fnsys.2013.00068

Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic Beta oscillations. *J. Neurosci.* 32, 1791–1802. doi: 10.1523/JNEUROSCI.4107-11.2012

Fukushima, H., Hirata, S., Ueno, A., Matsuda, G., Fuwa, K., Sugama, K., et al. (2010). Neural correlates of face and object perception in an awake chimpanzee (*Pan troglodytes*) examined by scalp-surface event-related potentials. *PLoS ONE* 5:e13366. doi: 10.1371/journal.pone.0013366

Gámez, J., Yc K, Ayala, Y. A., Dotov, D., Prado, L., and Merchant, H. (2018). Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Ann. N. Y. Acad. Sci.* doi: 10.1111/nyas.13671. [Epub ahead of print].

Gil-da-Costa, R., Stoner, G. R., Fung, R., and Albright, T. D. (2013). Nonhuman primate model of schizophrenia using a noninvasive EEG method. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15425–15430. doi: 10.1073/pnas.1312264110

Godlove, D. C., Emeric, E. E., Segovis, C. M., Young, M. S., Schall, J. D., and Woodman, G. F. (2011). Event-related potentials elicited by errors during the stop-signal task. I. Macaque monkeys. *J. Neurosci.* 31, 15640–15649. doi: 10.1523/JNEUROSCI.3349-11.2011

Grahn, J. A., and Brett, M. (2009). Impairment of beat-based rhythm discrimination in Parkinson's disease. *Cortex* 45, 54–61. doi: 10.1016/j.cortex.2008.01.005

Harms, L., Michie, P. T., and Näätänen, R. (2016). Criteria for determining whether mismatch responses exist in animal models: focus on rodents. *Biol. Psychol.* 116, 28–35. doi: 10.1016/j.biopsycho.2015.07.006

Hasegawa, A., Okanoya, K., Hasegawa, T., and Seki, Y. (2011). Rhythmic synchronization tapping to an audio–visual metronome in budgerigars. *Sci. Rep.* 1:120. doi: 10.1038/srep00120

Hattori, Y., Tomonaga, M., and Matsuzawa, T. (2015). Distractor effect of auditory rhythms on self-paced tapping in chimpanzees and humans. *PLoS ONE* 10:e0130682. doi: 10.1371/journal.pone.0130682

Hirata, S., Matsuda, G., Ueno, A., Fukushima, H., Fuwa, K., Sugama, K., et al. (2013). Brain response to affective pictures in the chimpanzee. *Sci. Rep.* 3:1342. doi: 10.1038/srep01342

Hoeschele, M., Merchant, H., Kikuchi, Y., Hattori, Y., and Ten Cate, C. (2015). Searching for the origins of musicality across species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140094. doi: 10.1098/rstb.2014.0094

Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Ann. N. Y. Acad. Sci.* 1252, 85–91. doi: 10.1111/j.1749-6632.2011.06402.x

Honing, H. (2018a). "Musicality as an Upbeat to Music: Introduction and Research Agenda," in *The Origins of Musicality*, ed H. Honing (Cambridge, MA: MIT Press ), 3–20.

Honing, H. (Ed.). (2018b). *The Origins of Musicality*. Cambridge, MA: MIT Press.

Honing, H., Bouwer, F. L., and Háden, G. P. (2014). "Perceiving temporal regularity in music: the role of auditory event-related potentials (ERPs) in probing beat perception," in *Neurobiology of interval timing*, eds V. Merchant, and H. de Lafuente (Berlin; Heidelberg: Springer ), 305–323.

Honing, H., and Merchant, H. (2014). Differences in auditory timing between human and non-human primates. *Behav. Brain Sci.* 27, 557–558. doi: 10.1017/S0140525X13004056

Honing, H., Merchant, H., Háden, G. P., Prado, L., and Bartolo, R. (2012). Rhesus monkeys (*Macaca mulatta*) detect rhythmic groups in music, but not the beat. *PLoS ONE* 7:e51369. doi: 10.1371/journal.pone.0051369

Honing, H., and Ploeger, A. (2012). Cognition and the evolution of music: pitfalls and prospects. *Top. Cogn. Sci.* 4, 513–524. doi: 10.1111/j.1756-8765.2012.01210.x

Honing, H., ten Cate, C., Peretz, I., and Trehub, S. E. (2015). Without it no music: cognition, biology and evolution of musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140088. doi: 10.1098/rstb.2014.0088

Howell, T. J., Conduit, R., Toukhsati, S., and Bennett, P. (2012). Auditory stimulus discrimination recorded in dogs, as indicated by mismatch negativity (MMN). *Behav. Process.* 89, 8–13. doi: 10.1016/j.beproc.2011.09.009

Itoh, K., Nejime, M., Konoike, N., Nakada, T., and Nakamura, K. (2015). Noninvasive scalp recording of cortical auditory evoked potentials in the alert macaque monkey. *Hear. Res.* 327, 117–125. doi: 10.1016/j.heares.2015.05.007

Javitt, D. C., Jayachandra, M., Lindsley, R. W., Specht, C. M., and Schroeder, C. E. (2000). Schizophrenia-like deficits in auditory P1 and N1 refractoriness induced by the psychomimetic agent phencyclidine (PCP). *Clin. Neurophysiol.* 111, 833–836. doi: 10.1016/S1388-2457(99)00313-2

Javitt, D. C., Steinschneider, M., Schroeder, C. E., Vaughan H. G. Jr., and Arezzo, J. C. (1994). Detection of stimulus deviance within primate primary auditory cortex: intracortical mechanisms of mismatch negativity (MMN) generation. *Brain Res.* 667, 192–200. doi: 10.1016/0006-8993(94)91496-6

Kimura, M. (1992). Behavioral modulation of sensory responses of primate putamen neurons. *Brain Res.* 578, 204–214.

Ladinig, O., Honing, H., Háden, G. P., and Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Percept.* 26, 377–386. doi: 10.1525/mp.2011.28.4.444

Ladinig, O., Honing, H., Háden, G. P., and Winkler, I. (2011). Erratum to Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Percept.* 26, 377–386. doi: 10.1525/mp.2009.26.4.377

Large, E. W., and Gray, P. M. (2015). Spontaneous tempo and rhythmic entrainment in a Bonobo. *J. Comp. Psychol.* 129, 317–328. doi: 10.1037/com0000011

Laughlin, N. K., Hartup, B. K., Lasky, R. E., Meier, M. M., and Hecox, K. E. (1999). The development of auditory event related potentials in the rhesus monkey (Macaca mulatta). *Dev. Psychobiol.* 34, 37–56.

London, J. (2002). Cognitive constraints on metric systems: some observations and hypotheses. *Music Percept.* 19, 529–550. doi: 10.1525/mp.2002.19.4.529

Mathias, B., Lidji, P., Honing, H., Palmer, C., and Peretz, I. (2016). Electrical brain responses to beat irregularities in two cases of beat deafness. *Front. Neurosci.* 10:40. doi: 10.3389/fnins.2016.00040

Mendoza, G., and Merchant, H. (2014). Motor system evolution and the emergence of high cognitive functions. *Prog. Neurobiol.* 122, 73–93. doi: 10.1016/j.pneurobio.2014.09.001

Merchant, H., and Bartolo, R. (2017). *Primate beta oscillations and rhythmic behaviors. J. Neural Transm.* 125, 461–470. doi: 10.1007/s00702-017-1716-9

Merchant, H., Battaglia-Mayer, A., and Georgopoulos, A. P. (2004). Neural responses during interception of real and apparent circularly moving stimuli in motor cortex and area 7a. *Cereb. Cortex* 14, 314–331. doi: 10.1093/cercor/bhg130

Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., and Fitch, W. T. (2015a). Finding the beat : a neural perspective across humans and non-human primates. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140093. doi: 10.1098/rstb.2014.0093

Merchant, H., and Honing, H. (2014). Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front. Neurosci.* 7:274. doi: 10.3389/fnins.2013.00274

Merchant, H., Pérez, O., Bartolo, R., Méndez, J. C., Mendoza, G., Gámez, J., et al. (2015b). Sensorimotor neural dynamics during isochronous tapping in the medial premotor cortex of the macaque. *Eur. J. Neurosci.* 41, 586–602. doi: 10.1111/ejn.12811

Merchant, H., and Yarrow, K. (2016). How the motor system both encodes and influences our sense of time. *Curr. Opin. Behav. Sci.* 8, 22–27. doi: 10.1016/j.cobeha.2016.01.006

Merchant, H., Zarco, W., Pérez, O., Prado, L., and Bartolo, R. (2011). Measuring time with different neural chronometers during a synchronization-continuation task. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19784–19789. doi: 10.1073/pnas.1112933108

Morillon, B., Schroeder, C. E., and Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nat. Commun.* 5:5255. doi: 10.1038/ncomms6255

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Nelken, I., and Ulanovsky, N. (2007). Mismatch negativity and stimulus-specific adaptation in animal models. *J. Psychophysiol.* 21, 214–223. doi: 10.1027/0269-8803.21.34.214

Patel, A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Percept.* 24, 99–104. doi: 10.1525/mp.2006.24.1.99

Patel, A. D. (2014). The evolutionary biology of musical rhythm: was Darwin wrong? *PLoS Biol.* 12:e1001821. doi: 10.1371/journal.pbio.1001821

Patel, A. D. (2018). "Music as a Transformative Technology of the mind: An Update," in *The Origins of Musicality,* ed H. Honing (Cambridge, MA: MIT Press), 113–126.

Patel, A. D., and Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front. Syst. Neurosci.* 8:57. doi: 10.3389/fnsys.2014.00057

Patel, A. D., Iversen, J. R., Bregman, M. R., and Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Curr. Biol.* 19, 827–830. doi: 10.1016/j.cub.2009.03.038

Pereira, D. R., Cardoso, S., Ferreira-Santos, F., Fernandes, C., Cunha-Reis, C., Paiva, T. O., et al. (2014). Effects of inter-stimulus interval (ISI) duration on the N1 and P2 components of the auditory event-related potential. *Int. J. Psychophysiol.* 94, 311–318. doi: 10.1016/j.ijpsycho.2014.09.012

Peretz, I. (2016). Neurobiology of congenital Amusia. *Trends Cogn. Sci.* 20, 857–867. doi: 10.1016/j.tics.2016.09.002

Petkov, C. I., and Jarvis, E. D. (2014). The basal ganglia within a cognitive system in birds and mammals. *Behav. Brain Sci.* 37, 568–569. doi: 10.1017/S0140525X13004160

Picton, T. W., Hillyard, S. A., Krausz, H. I., and Galambos, R. (1974). Human auditory evoked potentials. I: evaluation of components. *Electroencephalogr. Clin. Neurophysiol.* 36, 179–190.

Pincze, Z., Lakatos, P., Rajkai, C., Ulbert, I., and Karmos, G. (2001). Separation of mismatch negativity and the N1 in the auditory cortex of the cat: a topograhic study. *Clin. Neurophysiol.* 198, 778–784. doi: 10.1016/S1388-2457(01)00509-0

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331

Reinhart, R. M., Heitz, R. P., Purcell, B. A, Weigand, P. K., Schall, J. D., and Woodman, G. F. (2012). Homologous mechanisms of visuospatial working memory maintenance in macaque and human: properties and sources. *J. Neurosci.* 32, 7711–7722. doi: 10.1523/JNEUROSCI.0215-12.2012

Ross, J. M., Iversen, J. R., and Balasubramaniam, R. (2017). The role of posterior parietal cortex in beat-based timing perception: a continuous theta burst stimulation study. *J. Cogn. Neurosci.* 30, 634–643. doi: 10.1162/jocn_a_01237

Schachner, A., Brady, T. F., Pepperberg, I. M., and Hauser, M. D. (2009). Spontaneous motor entrainment to music in multiple vocal mimicking species. *Curr. Biol.* 19, 831–836. doi: 10.1016/j.cub.2009.03.061

Schall, U., Müller, B. W., Kärgel, C., and Güntürkün, O. (2015). Electrophysiological mismatch response recorded in awake pigeons from the avian functional equivalent of the primary auditory cortex. *Neuroreport* 26, 239–244. doi: 10.1097/WNR.0000000000000323

Schwartze, M., Farrugia, N., and Kotz, S. A. (2013). Dissociation of formal and temporal predictability in early auditory evoked potentials. *Neuropsychologia* 51, 320–325. doi: 10.1016/j.neuropsychologia.2012.09.037

Schwartze, M., and Kotz, S. A. (2015). The timing of regular sequences: production, perception, and covariation. *J. Cogn. Neurosci.* 27, 1697–1707. doi: 10.1162/jocn_a_00805

Schwartze, M., Rothermich, K., Schmidt-Kassow, M., and Kotz, S. A. (2011). Temporal regularity effects on pre-attentive and attentive processing of deviance. *Biol. Psychol.* 87, 146–151. doi: 10.1016/j.biopsycho.2011.02.021

Selezneva, E., Deike, S., Knyazeva, S., Scheich, H., Brechmann, A., and Brosch, M. (2013). Rhythm sensitivity in macaque monkeys. *Front. Syst. Neurosci.* 7:49. doi: 10.3389/fnsys.2013.00049

Takeya, R., Kameda, M., Patel, A. D., and Tanaka, M. (2017). Predictive and tempo-flexible synchronization to a visual metronome in monkeys. *Sci Rep.* 7, 1–12. doi: 10.1038/s41598-017-06417-3

Teichert, T. (2016). Tonal frequency affects amplitude but not topography of rhesus monkey cranial EEG components. *Hear. Res.* 336, 29–43. doi: 10.1016/j.heares.2016.04.001

Teki, S., and Griffiths, T. D.(2016). Brain bases of working memory for time intervals in rhythmic sequences. *Front. Neurosci.* 10:239. doi: 10.3389/fnins.2016.00239

Teki, S., Grube, M., Kumar, S., and Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *J. Neurosci.* 31, 3805–3812. doi: 10.1523/JNEUROSCI.5561-10.2011

Trehub, S. E. (2003). The developmental origins of musicality. *Nat. Neurosci.* 6, 669–673. doi: 10.1038/nn1084

Trehub, S. E., Becker, J., and Morley, I. (2015). Cross-cultural perspectives on music and musicality. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140096. doi: 10.1098/rstb.2014.0096

Ueno, A., Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., Matsuda, G., et al. (2008). Auditory ERPs to stimulus deviance in an awake chimpanzee (Pan troglodytes): towards hominid cognitive neurosciences. *PLoS ONE* 3:e1442. doi: 10.1371/journal.pone.0001442

Ueno, A., Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., Matsuda, G., et al. (2010). Brain activity in an awake chimpanzee in response to the sound of her own name. *Biol. Lett.* 6, 311–313. doi: 10.1098/rsbl.2009.0864

Wilson, M., and Cook, P. F. (2016). *Rhythmic entrainment: Why humans want to, fireflies can't help it, pet* birds try, and sea lions have to be bribed. *Psychon. Bull. Rev*. 23, 1647–1659. doi: 10.3758/s13423-016-1013-x

Winkler, I. (2007). Interpreting the mismatch negativity. *J. Psychophysiol.* 21, 147–163. doi: 10.1027/0269-8803.21.34.147

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., and Honing, H. (2009). Newborn infants detect the beat in music. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2468–2471. doi: 10.1073/pnas.0809035106

Zarco, W., Merchant, H., Prado, L., and Mendez, J. C. (2009). Subsecond timing in primates: comparison of interval production between human subjects and rhesus monkeys. *J. Neurophysiol.* 102, 3191–3202. doi: 10.1152/jn.00066.2009

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2018 Honing, Bouwer, Prado and Merchant. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 04 October 2018
doi: 10.3389/fpsyg.2018.01680

# The Human Nature of Music

*Stephen Malloch*[1,2]* *and Colwyn Trevarthen*[3]*

[1] Westmead Psychotherapy Program, Sydney Medical School, The University of Sydney, Sydney, NSW, Australia,
[2] The MARCS Institute for Brain, Behaviour, and Development, Western Sydney University, Sydney, NSW, Australia,
[3] Department of Psychology, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, Edinburgh, United Kingdom

Music is at the centre of what it means to be human – it is the sounds of human bodies and minds moving in creative, story-making ways. We argue that music comes from the way in which knowing bodies (Merleau-Ponty) prospectively explore the environment using habitual 'patterns of action,' which we have identified as our innate 'communicative musicality.' To support our argument, we present short case studies of infant interactions using micro analyses of video and audio recordings to show the timings and shapes of intersubjective vocalizations and body movements of adult and child while they improvise shared narratives of meaning. Following a survey of the history of discoveries of infant abilities, we propose that the gestural narrative structures of voice and body seen as infants communicate with loving caregivers are the building blocks of what become particular cultural instances of the art of music, and of dance, theatre and other temporal arts. Children enter into a musical culture where their innate communicative musicality can be encouraged and strengthened through sensitive, respectful, playful, culturally informed teaching in companionship. The central importance of our abilities for music as part of what sustains our well-being is supported by evidence that communicative musicality strengthens emotions of social resilience to aid recovery from mental stress and illness. Drawing on the experience of the first author as a counsellor, we argue that the strength of one person's communicative musicality can support the vitality of another's through the application of skilful techniques that encourage an intimate, supportive, therapeutic, spirited companionship. Turning to brain science, we focus on hemispheric differences and the affective neuroscience of Jaak Panksepp. We emphasize that the psychobiological purpose of our innate musicality grows from the integrated rhythms of energy in the brain for prospective, sensation-seeking affective guidance of vitality of movement. We conclude with a Coda that recalls the philosophy of the Scottish Enlightenment, which built on the work of Heraclitus and Spinoza. This view places the shared experience of sensations of living – our communicative musicality – as inspiration for rules of logic formulated in symbols of language.

**Keywords: musicking, motor intelligence, gestural narrative, infant musicality, cultural learning**

OPEN ACCESS

***Edited by:***
*Aleksey Nikolsky,*
*Braavo! Enterprises, United States*

***Reviewed by:***
*Antonio Damasio,*
*University of Southern California, United States*
*Marc Leman,*
*Ghent University, Belgium*
*Jerome Lewis,*
*University College London, United Kingdom*

***Correspondence:**
*Stephen Malloch*
*stephen.malloch@heartmind.com.au*
*Colwyn Trevarthen*
*c.trevarthen@ed.ac.uk*

***Specialty section:***
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

***Received:*** *18 February 2018*
***Accepted:*** *21 August 2018*
***Published:*** *04 October 2018*

***Citation:***
*Malloch S and Trevarthen C (2018) The Human Nature of Music. Front. Psychol. 9:1680. doi: 10.3389/fpsyg.2018.01680*

> "There are certain aspects of the so-called 'inner life'—physical or mental —which have formal properties similar to those of music—patterns of motion and rest, of tension and release, of agreement and disagreement, preparation, fulfilment, excitation, sudden change, etc. Langer (1942, p. 228).

> "The function of music is to enhance in some way the quality of individual experience and human relationships; its structures are reflections of patterns of human relations, and the value of a piece of music as music is inseparable from its value as an expression of human experience" Blacking (1995, p.31).

> "The act of musicking establishes in the place where it is happening a set of relationships, and it is in those relationships that the meaning of the act lies. They are to be found not only between those organized sounds which are conventionally thought of as being the stuff of musical meaning but also between the people who are taking part, in whatever capacity, in the performance" Small (1998, p.9).

## PRELUDE

We present a view that places our ability to create and appreciate music at the center of what it means to be human. We argue that music is the sounds of human bodies, voices and minds – our personalities – moving in creative, story-making ways. These stories, which we want to share and listen to, are born from awareness of a complex body evolved for moving with an imaginative, future seeking mind in collaboration with other human bodies and minds. Musical stories do not need words for the creation of rich and inspiring narratives of meaning.

We adopt the word 'musicking' (as used above by Christopher Small) to draw attention to the embodied energy that creates music, and which moves us, emotionally and bodily. Further, we argue that music comes from the way in which *knowing bodies* (Merleau-Ponty, 2012 [1945], p. 431) prospectively explore the environment using habitual 'patterns of action,' which we have identified as our innate 'communicative musicality,' observed while infants are in intimate communication with loving caregivers (Malloch and Trevarthen, 2009a). In short case studies of infant interactions with micro analyses of video and audio recordings, we show communicative musicality in the timings and shapes of intersubjective vocalizations and body movements of adult and child that improvise with delight shared narratives of meaning.

Following a survey of the history of discoveries of infant abilities, we propose that the gestural narrative structures of voice and body seen as infants communicate with loving caregivers, 'protonarrative envelopes' of expression for ideas of activity (Panksepp and Bernatzky, 2002; Panksepp and Trevarthen, 2009), are the building blocks of what become particular cultural instances of the art of music, and of dance, theater and other temporal arts (Blacking, 1995).

As the child grows and becomes a toddler, she or he eagerly takes part in a children's musical culture of the playground (Bjørkvold, 1992). Soon more formal education with a teacher leads the way to the learning of traditional musical techniques. It is at this point that the child's innate body vitality of communicative musicality can be encouraged and strengthened through sensitive, respectful, playful, culturally informed teaching (Ingold, 2018). On the other hand, it may wither under the weight of enforced discipline for the sake of conforming to pre-existing cultural rules without attention to the initiative and pleasure of the learner's own music-making.

The central importance of our abilities for music as part of what sustains our well-being is supported by evidence that communicative musicality strengthens emotions of social resilience to recover from mental stress and illness (Pavlicevic, 1997, 1999, 2000). Drawing on the experience of the first author as a counselor, we argue that the strength of one person's communicative musicality can support the vitality of another's through the application of skilful techniques that encourage an intimate, supportive, therapeutic, spirited companionship.

Turning to brain science, focussing on hemispheric differences in performance and in response to music, and the affective neuroscience of Jaak Panksepp (1998), we emphasize that the psychobiological purpose of our 'muse within' (Bjørkvold, 1992) grows from the integrated rhythms of neural energy for prospective, sensation-seeking affective guidance of vitality of movement in the brain (Goodrich, 2010; Stern, 2010).

We conclude with a Coda – an enquiry into the philosophy of the Scottish Enlightenment, which built on the work of philosophers Heraclitus and Spinoza. This view of living in community gives innate sympathy or 'feeling with' other humans a fundamental role within a duplex mind seeking harmony in relationships by attunement of motives (Hutcheson, 1729, 1755). It places the shared experience of sensations of living – our communicative musicality – ahead of logic formulated in symbols of language.

## MUSIC MOVES US – EMBODIED NARRATIVES OF MOVEMENT

Small (1998) calls attention to music as intention in activity by using the verb *musicking* – participating as performer or listener with attention to the sounds created and the appreciation and participation by others. The compelling quality of music comes from the relationships of sounds, bodies and psyches. 'Musicking' points to our musical life in active 'I-Thou' relationships. Only in this intimacy of consciousness and its interests can we share 'I-It' identification and use of objects, giving things we use, including musical compositions, meaning (Buber, 1923/1970).

'I-Thou' relationships are entered into through the body. The philosopher Merleau-Ponty (2012 [1945]) writes that "The subject only achieves his ipseity [individual personality, selfhood] by actually being a body, and by entering into the world through his body. . . The ontological world and body that we uncover at the core of the subject are not the world and the body as *ideas*; rather, they are the world itself condensed into a comprehensive whole and the body itself as a *knowing-body*" (p. 431; italics added). Musicking is *knowing bodies* coming alive in the sounds they make. Scores and other tools that record the product of musicking, performed or imagined, aid the retention of ideas, as semantics of language does, and they serve discussion and analysis – but they are not the same as the breathing, moving, *embodied* experience of human musicking (Mithen, 2005).

Musicking is the expression of the sensations of what we call our *communicative musicality*, for the purpose of creating music

that is enlivening and 'beautiful' (Malloch, 1999; Malloch and Trevarthen, 2009a; Trevarthen, 2015; Trevarthen and Malloch, 2017b). We define communicative musicality as our innate skill for moving, remembering and planning projects in sympathy with others through time, creating an endless variety of dramatic temporal narratives in song or instrumental music. We describe this life-sharing in movement as having three components:

*Pulse* – a regular succession through time of discrete movements (which may, for example, be used to create sound for music, or to create movement with music – dance) using our felt sense of acting which enables the 'future-creating' predictive process by which a person may anticipate or create what happens next and when.

*Quality* – consisting of the contours of expressive vocal and body gesture, shaping our felt sense of time in movement. These contours can consist of psychoacoustic attributes of vocalizations – timbre, pitch, volume – or attributes of direction and intensity of the moving body perceived in any modality.

*Narratives* of individual experience and of companionship, built from sequences of co-created gestures which have particular attributes of *pulse* and *quality* that bring aesthetic pleasure (Malloch and Trevarthen, 2009b; Trevarthen and Malloch, 2017b).

With music we create memorable poetic events in signs that express in sound our experience of living together in the creating vitality of 'the present moment' (Stern, 2004, 2010). The anthropologist and ethnologist Claude Lévi-Strauss draws attention of linguists to the structured 'raw' emotive power of music (beyond what words may be 'cooked' to say).

> "In the first volume of the *Mythologiques, Le Cru et le Cuit.* Lévi-Strauss refers to music as a unique system of signs possessing 'its own peculiar vehicle which does not admit of any general, extramusical use'. Yet he also allows that music has levels of structure analogous to the phonemes and sentences of language. The absence of words as the connecting level is an obvious and pertinent fact in the structuring of meaning within music as a sign system." (Champagne, 1990, p. 76).

Goodrich (2010), in her appreciation of the contribution of neuroscientists Llinás (2001) and Buzsáki (2006) to the science of the mind for skilled movement, cites Llinás' evidence on the role of intuitive structural 'rules,' seen also in a musical performance.

> "Llinás describes another method of keeping movement as efficient as possible: motor 'Fixed Action Patterns' (FAPs), distinct and complicated 'habits' of movement built from reflexes, habits that we develop to streamline both neural action and muscle movement. These are not entirely fixed, despite their name; they are constantly undergoing modification, adaptation, refinement, and they overlap each other... Llinás even argues that the extraordinarily precise motor control of Jascha Heifetz playing Tchaikovsky's violin concerto is composed of highly elaborated and refined FAPs, a description most instrumentalists would find absolutely plausible" (Goodrich, 2010, p. 339).

As Llinás himself writes,

> "Can playing a violin concerto be a FAP? Well, not all of it, but a large portion. Indeed, the unique and at once recognisable style of play Mr. Heifetz brings to the instrument is a FAP, enriched and modulated by the specifics of the concert, generated by the voluntary motor system" (Llinás, p. 136).

We add that skilled FAPs are not "composed of reflexes" as separate automatic responses. Rather they are purposeful projects that are animated to be developed imaginatively, and affectively, with exploration of their biomechanical "degrees of freedom," as in Nikolai Bernstein's detailed description of how a toddler learns to become a virtuoso in bipedal locomotion, which he calls *The Genesis of the Biodynamical Structure of the Locomotor Act* (Bernstein, 1967, p. 78). The testing of these locomotor acts is with an immediate and essential estimation by *gut feelings* (Porges, 2011) of any risks or benefits, any fears or joy, they may entail within the body.

## MUSIC REFLECTS THE FELT-SENSE OF OUR FUTURE-EXPLORING MOTOR INTELLIGENCE

Consciousness is created as the ongoing sense of self-in-movement with which we experience and manipulate the world around us. Its origin is in our evolutionary animal past, evolved for new collaborative, creative projects, regulated between us by affective expressions of feelings of vitality from within our bodies (Sherrington, 1955; Panksepp, 1998; Damasio, 2003; Mithen, 2005; Stern, 2010; Eisenberg and Sulik, 2012).

Using the philosopher and psychologist James (1892/1985) as a starting point for exploring the intimacy of feeling that supports and guides psychotherapy, Russell Meares (2005, p.18), following the 'conversational model' of therapy developed with his collaborator psychiatrist/psychotherapist Hobson (1985), identifies five dimensions of the self:

1. *awareness* which is necessary for the experience;
2. there is a *shape* to this inner life;
3. there is *a sense of its ongoingness*;
4. our inner life has a *connectedness or unity*;
5. our experience of our inner life goes on inside a virtual container which is *our background emotional state* and the *background experience of our body*.

These sensuous qualities of the experienced self are expressed in music, and in other temporal arts, as 'the human seriousness of play' (Turner, 1982). Music, as Susan Langer says so clearly in the quote at the start of this paper, has qualities of this inner life described by Meares as shape, ongoingness and flow, connectedness and unity. The notion of music as expressive of the movements of our inner life has also been explored by music theorists, most notably Ernst Kurth (1991). Likewise, in his book *Self comes to Mind*, Antonio Damasio likens all our emotion and feeling to a 'musical score' that accompanies other ongoing mental process (Damasio, 2010, p.254).

The ultimate motivation for creating music can even be traced to the cellular level. In *Man On His Nature* (Sherrington, 1955), in a chapter entitled *The Wisdom of the Body,* the creator of modern neurophysiology Charles Sherrington called the coming together of communities of cells into the integrated body,

nervous system and brain of a person "an act of imagination" (p. 103). Neuroscientist Rudolfo Llinás also grants subjectivity, a sense of self, to all forms of life. "Irritability [i.e., responding to external stimuli with organized, goal-directed behavior] and subjectivity, in a very primitive sense, are properties originally belonging to single cells" (Llinás, 2001, p.113). "Thinking", writes Llinás (2001, p.62), "ultimately represents movement, not just of body parts or objects in the external world, but of perceptions and complex ideas as well."

Intrinsic to the sociability of this intelligence of movement is sensitivity for the exploration of the future, which is woven into our creation and experience of music. Karl Lashley (1951) reflecting on the evolution of animal movement, proposed that the ability to predict what might come next, and to plan the 'serial ordering' of separate actions, may be understood as the foundation for our logical reasoning as an individual, as it is for the grammar or syntax and prosody of our communication in language. It is essential for musicking. A restless future-seeking intelligence, with our urge to share it, inspires us to express our personalities as 'story-telling creatures,' who want to share, and evaluate, others' stories (Bruner, 1996, 2003).

All animal life depends on motivated movement – the urge to explore with curiosity – to move towards food with anticipation, to move away from a predator with fear, to interact playfully with a trusted friend (Eibl-Eibesfeldt, 1989; Panksepp and Biven, 2012; Bateson and Martin, 2013). A great achievement of modern science of the mind was the discovery by a young Russian psychologist Nikolai Bernstein of how all consciously made body movements depend upon an 'image of the future' (Bernstein, 1966, 1967).

Bernstein applied the new technology of movie photography to make refined 'cyclographic' diagrams of displacements of body parts, from which he analyzed the forces involved to fractions of a second. His findings reported in *Coordination and Regulation of Movements* became widely known in English translation in 1967, at the same time as video records of infant behaviors were described more accurately (see next Section The Genesis of Music in Infancy – A Short History of Discoveries), revealing their anticipatory motor control adapted for intelligent understanding of how objects may be manipulated, as well as for communication and cooperation (Trevarthen, 1984b, 1990b).

Our musical creativity and pleasure come from the way our body hopes to move, with rhythms and feelings of grace and biological 'knowing' (see Merleau-Ponty). The predicting, embodied self of a human being experiences time, force, space, movement, and intention/directionality in being. Together, these form the Gestalt of 'vitality' (Stern, 2010), the 'forms of feeling' (Hobson, 1985) by which we sense in ourselves and in others that movement, be that movement of the body or of a piece of music, is 'well-done' (Trevarthen and Malloch, 2017b).

## THE GENESIS OF MUSIC IN INFANCY – A SHORT HISTORY OF DISCOVERIES

The ability to create meaning with others through wordless structured gestural narratives, that is, our communicative musicality, emerges from before birth and in infancy. From this innate musicality come the various cultural forms of music.

Any attempt to understand how human life has evolved its unique cultural habits needs to start with observing what infants know and can do. Organisms regulate the development of their lives by growing structures and processes from within their vitality, by *autopoiesis* that requires anticipation of adaptive functions. And they must develop and protect their abilities in response to environmental affordances and dangers, with *consensuality* (Maturana and Varela, 1980; Maturana et al., 1995). Infants are ready for human cultural invention and collaboration as newly hatched birds are ready for flying – within 'the biology of love' (Maturana and Varela, 1980; Maturana and Verden-Zoller, 2008). All organisms reach out in time and space to make use of the 'affordances' for thought and action (Gibson, 1979).

Infants have no language to learn what other humans know, or what ancestors knew. But the vitality of their spontaneous communicative musicality, highly coordinated and adapted to be shared through narratives with sympathetic and playful companions, enables meaningful communication in the 'present moment' (Stern, 2004; **Figure 1**, Upper Right), which may build serviceable memories extended in space and time (Donaldson, 1992).

In this section we review changes in understanding of infant abilities over the past century which can help explain the peculiar way music has in the past been seen by some leading psychologists and linguists as a relatively insignificant epiphenomenon, learned for play, not, as we argue in this paper, a source of all talents for communication, rooted in our innate human communicative musicality of knowing bodies (Merleau-Ponty, 2012 [1945]) moving with prospective intuition to engage the world in company, 'intersubjectively,' from the start (Zeedyk, 2006).

Two leading scholars in medical science and the science of child development in the past century, Freud (1923) and Piaget (1958, 1966), declared that infants must be born without conscious selves conceiving an external world, and unable to adapt their movements to the expressive behavior of other people. The playful and emotionally charged behaviors of mothers and other affectionate carers were considered inessential to the young infant, who needed only responses to reflex demands for food, comfort and sleep.

Then René Spitz (1945) and Bowlby (1958) revealed the devastating effects on a child's emotional well-being of separation from maternal care in routine hospital care with nursing directed only to respond to those reflex demands. Spitz observed that babies develop smiling between 2 and 5 months to regulate social contacts (Spitz and Wolf, 1946), and he went on to study the independent will of the baby to regulate engagements of care or communication, by nodding the head for 'yes' or shaking for 'no' (Spitz, 1957).

In the 1960s a major shift in understanding of the creative mental abilities of infants was inspired by a project of the educational psychologist Bruner (1966, 1968), and the pediatrician Brazelton (1961, 1979), who perceived that infants are gifted with sensibilities for imaginative play and ready to start cultural learning from the first weeks after birth. Supported by

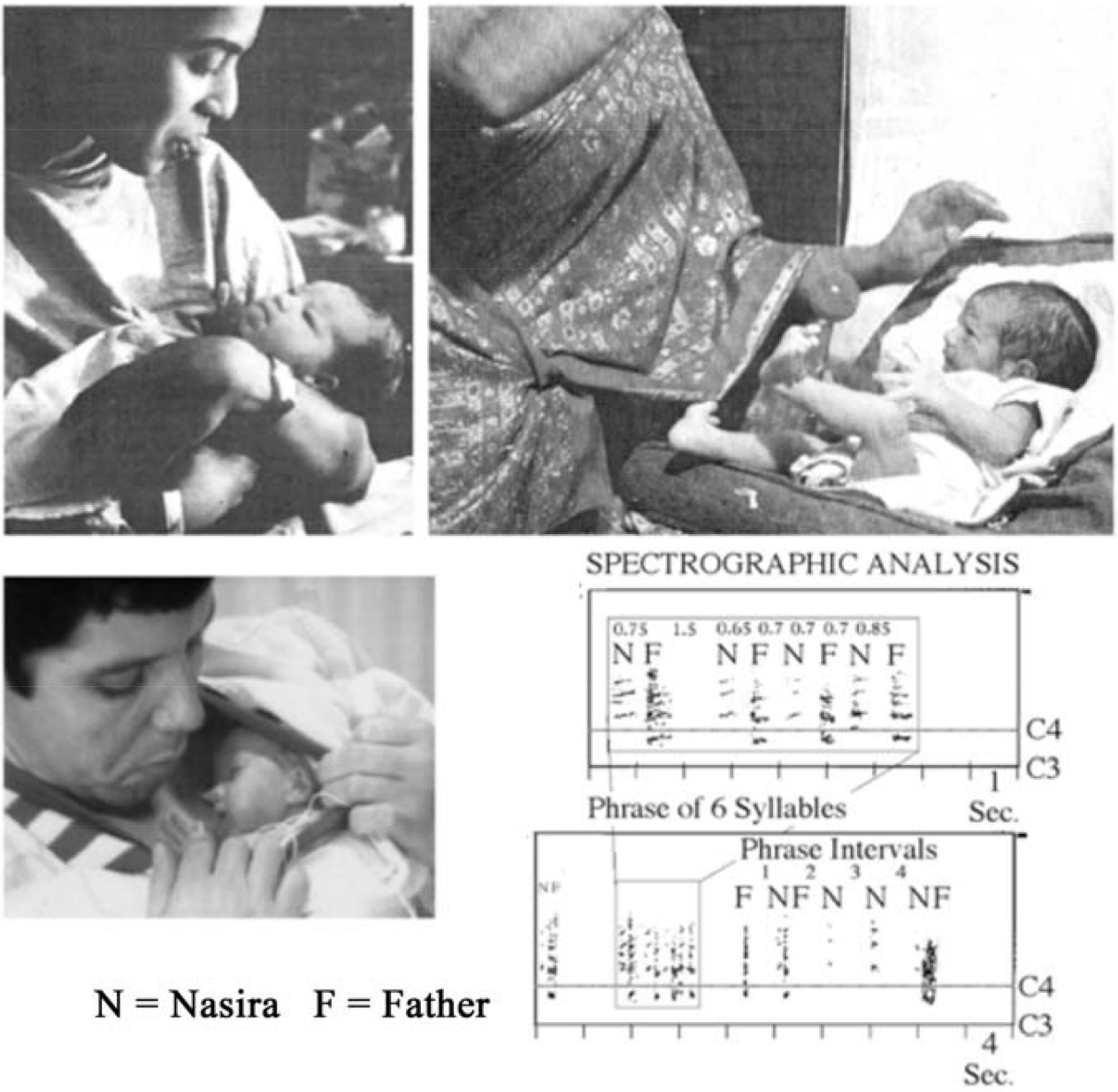



**FIGURE 1 |** Inborn musicality shared in movement. Upper Left: Infant less than one hour after birth watches her mother's tongue protrusion and imitates. Upper Right: On the day of birth, a baby in a hospital in India shares a game with a woman who moves a red ball. The baby tracks it with coordinated movements of her eyes and head, both hands and one foot. Lower: A two-month premature girl, Nasira, with her father, who holds her against his body in 'kangaroo' care. They exchange short 'coo' sounds with precisely shared rhythm. Upper Left and Upper Right: Photos for own use of second author from colleagues Vasudevi Reddy and Kevan Bundell. Part reproduced from Trevarthen (2015), Figure 1, p. 131. Lower: Photo from Trevarthen (2008), Figure 2, p. 22. Original spectrograph in Malloch (1999), Figure 3, p. 37.

insights of Charles Darwin and by new findings of anthropology and animal ethology they studied infant initiatives to perceive and use objects, and they were impressed by the intimate reciprocal imitation that develops between infants and affectionate parents and caregivers who offer playful collaboration with the child's rhythms and qualities of movement. Film studies showed that young infants make complex shifts of posture and hand gestures that are regulated rhythmically, similar to the same movements of adults (Bruner, 1968; Trevarthen, 1974).

While this new appreciation of infant abilities was developed at the Center for Cognitive Studies at Harvard, radically transforming the 'cognitive revolution' that was announced there by George Miller, Noam Chomsky and Jerome Bruner in 1960, nearby at the Massachusetts Institute of Technology, a project

initiated by Bullowa (1979) sought evidence on the behaviors that regulate dialog before language. Bullowa used information from anthropology to draw attention to the *measured dynamics* of communication.

> "For an infant to enter into the sharing of meaning he has to be in communication, which may be another way of saying *sharing rhythm*.... The problem is how two or more organisms can share innate biological rhythms in such a way as to achieve communication which can permit transmission of information they do not already share." (Bullowa, 1979, p. 15, italics added).

Wanting to understand how the rhythmic flow of dialog can be shared with a child too young for speech, she pointed a way to the appreciation of the role of human communicative musicality as inspiration for the development of behaviors for carrying meaning in language – thinking and communicating in words that will be acquired to specify facts, and to describe and think about how these facts are related or may be used.

The importance of rhythm and the graceful narratives of movement displayed by infants as they communicate purposes and feelings was revealed sixty years ago by a psychobiological approach using photography and movie film, then video. Discoveries were made that challenged the theory that infants had no minds, no sense of self, and therefore no sense of others (Zeedyk, 2006; Reddy, 2008). Most astonishing, and dismissed with derision by convinced rational mind-separate-from-body constructivists, was the finding that infants activate the many parts of their body with an exquisite sense of time, and that they can use the rhythms of expression skilfully to imitate in inter-synchrony with attentive responses from an adult (**Figure 1**).

In his work as a pediatrician, Brazelton (1961), developing his now famous Neonatal Assessment Scale (Brazelton, 1973; Brazelton and Nugent, 1995), accepted and encouraged the natural love mother and father felt for their new baby, and showed how appreciative the baby could be of their actions to each other and to the baby. This welcoming of the newborn as a person with intelligence and sociable impulses confirmed the parents' belief that they could communicate feelings and interests by responding to their baby's exquisitely timed looks, smiles, hand gestures and cooing with their own exquisitely timed gestures of voice and body. It transformed medical concern for the baby. As Brazelton declared in Margaret Bullowa's book, "The old model of thinking of the newborn infant as helpless and ready to be shaped by his environment prevented us from seeing his power as a communicant in the early mother-father-infant interaction. To see the neonate as chaotic or insensitive provided us with the capacity to see ourselves as acting 'on' rather than 'with' him" (Brazelton, 1979, p.79).

New attention to newborns within hours of their delivery, with the aid of films, led to confirmation that the baby could imitate adult expressions with careful timing of movements of eyes, face, mouth and hands (Maratos, 1973, 1982; Meltzoff and Moore, 1977; Kugiumutzakis, 1993; Nagy and Molnár, 1994, 2004; Nagy, 2011; Kugiumutzakis and Trevarthen, 2015; **Figure 1**, Upper Left). The findings proved that the baby is born with an *altero-ceptive awareness* of another person's body parts as having feelings in movement like their own *proprio-ceptive* ones. It also became clear that this consciousness appreciates the balance and drama of a collaborative narrative flow with shared rhythms – essential to our ability for musicking (Trevarthen, 1974, 1977, 2005a).

The pediatrician Sander, (1964, 1975; republished in Sander, 2008) recognized that an infant and caregiver create a coherent system of actions regulated with feelings of vitality in shared time. This dynamic collaboration was also discovered by Daniel Stern when he examined recordings of a mother playing with her three-month-old twins (Stern, 1971).

A stimulating contribution to this new approach came from the work of anthropologist and linguist Mary Catherine Bateson, daughter of anthropologists Gregory Bateson and Margaret Meade. In 1969 Bateson had her first child after beginning postgraduate studies at MIT with Margaret Bullowa, researching language development using statistical analysis of vocal expressions. Observing a film in Bullowa's collection as well as the experience of rich exchanges with her own infant opened her awareness of the form and timing of communication that developed in the first 3 months, which she called 'proto-conversation' (Bateson, 1971). She benefitted from attention to the field studies of Albert Scheflen on the stream of conversation (Scheflen, 1972) and Ray Birdwhistell on body movements in natural conversation (Birdwhistell, 1970).

Reviewing her work in Bullowa's book, she said:

> "... the mother and infant were collaborating in a pattern of more or less alternating, non-overlapping vocalization, the mother speaking brief sentences and the infant responding with coos and murmurs, together producing a brief joint performance similar to conversation, which I called 'proto conversation'. The study of timing and sequencing showed that certainly the mother and probably the infant, in addition to conforming in general to a regular pattern, were acting to sustain it or to restore it when it faltered, waiting for the expected vocalization from the other and then after a pause resuming vocalization, as if to elicit a response that had not been forthcoming. These interactions were characterized by a sort of delighted, ritualized courtesy and more or less sustained attention and mutual gaze. Many of the vocalizations were of types not described in the acoustic literature on infancy, since they were very brief and faint, and yet were crucial parts of the jointly sustained performances." (Bateson, 1979, p. 65).

Bateson's work confirmed Bruner's realization that Noam Chomsky's hypothesis of a specific Language Acquisition Device (LAD) or innate ability of the child to construct syntax, the grammatical order of words, to formulate ideas (Chomsky, 1965), paid no heed to the vital importance of a complementary ability of a parent to encourage enrichment of reference in communication with a young child, a Language Acquisition Support System (LASS). As Bruner, expressing his psychology of education, put it "the LADD needs a LASS" (Bruner, 1983). Child and adult share rules of imagination for all kinds of movement, including spoken propositions.

A follower of Chomsky's theory of the evolution of language as reasoning, Stephen Pinker, in his perhaps overly ambitiously titled *How the Mind Works* (Pinker, 1997) claimed, "As far as biological cause and effect are concerned, music is useless." He

gave no attention to movement and time, the communication of infants, playful children, affectionate parents, the poetry of music, or Einstein's theory of his mathematical invention as "sensations of bodily movement" (Hadamard, 1945), thus misunderstanding the origins and purpose of rational discourse (see Sections Communicative Musicality and Resilience of the Human Spirit and Musical Affections of the Embodied Human Brain).

We now have evidence from many studies analyzing behaviors that demonstrate that infants show a rich spectrum of expressive movements of the upper parts of their bodies (Trevarthen, 1984a), not just the 'categorical emotions' identified by Paul Ekman (Ekman and Friesen, 1975), but the 'complex social emotions' that Damasio (1999, 2010) describes as regulators of well-being in intimate interpersonal relations, and expression of a moral personality in society – expressions of such feelings as embarrassment, shame, guilt, contempt, compassion, and admiration.

Most importantly, study of recordings reveal that modulations of timing, of rhythms, and of the flow of vitality forms shared with infants have the characteristics recognized as musical (Trevarthen, 1990a). These have been precisely defined by acoustic analysis of vocalizations of adult and infant in dialogs and games (Malloch et al., 1997; Malloch, 1999; Trevarthen, 1999; Trehub, 2003).

## CASE STUDIES OF INFANT MUSICALITY

We summarize here key findings related to the growth of musical abilities from studies of infant individuals that we have reported previously.

First there is evidence from a recording made by Saskia van Rees in an intensive care unit in Amsterdam (van Rees and de Leeuw, 1993) that rhythms corresponding to those of human locomotion are present in vocalizations of a premature infant which are precisely coordinated with simple vocal exchanges with a caring father (**Figure 1**, Lower).

The recording of a two-month premature girl with her father, who was holding her under his clothes against his body in 'kangaroo' care, shows that they exchange short 'coo' sounds, the father imitating her sounds, with precise timing based on a comfortable walking rhythm of *andante* – one step every 0.7 s. Father (F) and the baby Naseera (N) are equally precise in their timing, which also shows what a phonetician would recognize as a 'final lengthening' characteristic of a spoken phrase - when they are ready to stop the dialog the interval lengthens to 0.85 s. Following the shared phrase with its syllable-length durations, they exchange single sounds separated at 4 s intervals, the normal duration of a short spoken phrase. The recording supports our contention that even a prematurely born baby is skilled in sharing a musical pulse (Trevarthen, 2009).

A recording with a blind 5-month-old girl illustrates intermodal attunement between the heard melody of a mother's song and the proprioceptive feelings in the body of the baby of a gesturing left arm and hand (**Figure 2**, Upper). The human ability to sense the shape of a melody within the body is intrinsic to our enjoyment of music as human communication (Stern, 2010). Maria is totally blind and has never seen her hand. Maria and her mother were assisting in a project of Professor Gunilla Preisler in Stockholm to aid communication with blind and deaf infants.

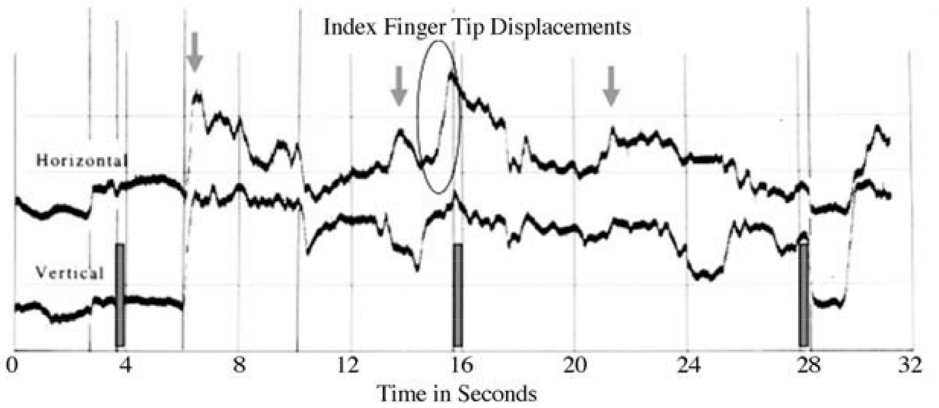
While her baby is lying down during bottle feeding, the mother sings two baby songs including "Mors Lille Olle," well-known throughout Scandinavia. It was not realized until later when the video was viewed that Maria was 'conducting' the melodies with delicate expressive movements of her left hand, while the right hand was making unrelated movements, stroking her body. When Professor of Music at Edinburgh University, Nigel Osborne, saw the film he said, "Yes she is conducting using the conventional movements of a professional conductor, describing a phrase with a sweeping movement, pointing up for a higher pitch, and dropping her wrist at the close of a verse – and she is making the movements with some anticipation." Microanalysis supported what he observed. At certain points in the course of the melody Maria's finger moves 300 milliseconds before the mother's voice. She knows the song well, and leads the 'performance' (Trevarthen, 1999; Schögler and Trevarthen, 2007).

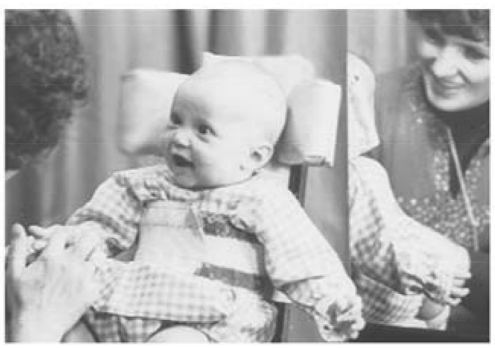
Although blind, Maria knows the feelings of anticipated movement of her hand, and uses them to sense and share the human vitality dynamics in her mother's voice. This *kinematic* sensibility was identified by Olga Maratos in her pioneering research in imitation as foundational for the ability of a young infant to reproduce another person's expression seen or heard (Maratos, 1973, 1982). Indeed, vocal perception, detecting the modulation of pitch and timing in an adult's voice sounds, develops much faster than vocal production. The infant may be tracking sound with reference to the kinesics of the fastest and most complex gestural movements of her hands.

When taking part in a nursery song, infants demonstrate sensitivity for melodic phrase structure, attending to the rhyming vowels at the ends of lines, and by 5 months the infant can vocalize a matching vowel in synchrony with the mother (Trevarthen, 2008). For our final example, the application of acoustic analysis with observations on gestural behaviors of infants in the middle of the first year, suggests that melodic patterns, common to different cultures, define four-line verses (Imberty, 2000), with a pattern of Introduction, Development, Climax and Resolution, identified in proto-conversations with two-month-olds (**Figure 2**, Lower, and **Figure 3**), and we note similarities with the sections recognized in classical Roman rhetoric or speech-making – *exordium, narratio, confutatio,* and *confirmatio*. In spite of very different conventions in musical performances in different communities, a parent, or a child, wanting to share the pleasure of songs and action games with a baby, naturally adopts the intuitive formula of a poetic verse to share a story of body movement.

Lastly, we present the work of Katerina Mazokopaki, a developmental psychologist who is a pianist and teacher of piano playing. She made a study of babies in Crete with her professor, Giannis Kugiumutzakis, an expert in analysis of imitative games with newborns (Mazokopaki and Kugiumutzakis, 2009). The babies were left alone in a familiar place at home amusing themselves. Then a recording of a Greek baby song came on. Between 3 and 10 months old they all reacted in the same way.





Round and round the gar-den,
• ‘ • ‘ • •
Ran a ted-dy **bear**,
• ‘ • ‘ • –
One step, two step,
• • • •
Tic-kl-y un-der **there.**
• ‘ ‘ • ‘ • –

Clappa, clappa handies
• ‘ • ‘ • •
Mommy’s at the **well**,
• ‘ • ‘ •
Daddy’s away to Hamilton
• ‘ • ‘ • ‘ •
To buy wee Emma a **bell.**
‘ • ‘ • ‘ •



**FIGURE 2 |** Communicating in the rhythm of narratives with baby songs. Upper: A five-month-old girl, who is totally blind, ‘conducts’ her mother’s singing with her left hand. The graph of her hand movements with her mother’s voice marks, with black bars, three moments in the narrative, at 4, 16, and 28 s, where the infant anticipates the mother’s voice by 0.3 s. These are times, 12 s apart, when lines of the slow lullaby commence. The arrows identify moments when there is perfect synchrony between the mother’s voice and the baby’s index finger. The circled index finger movement is 300 msec ahead of a lift in the pitch of the mother’s voice. Lower: Two infants share the poetry of action songs with their mothers. A four-month-old enjoys her mother’s singing of a song and accompanies the narrative with movements of her baby’s hand. And a six-month-old imitates clapping movements in synchrony with the song. Both songs show an iambic rhythm of short and long syllables, and rhyming vowels, ‘bear’ with ‘there’ and ‘well’ with ‘bell,’ which the babies imitate with their voices. (Upper: Analyses of video from Professor Gunilla Preisler, University of Stockholm. Previously published in Trevarthen (1999) and Schögler and Trevarthen (2007). Lower: From Trevarthen (2015), modified from Figure 4, p.137.)

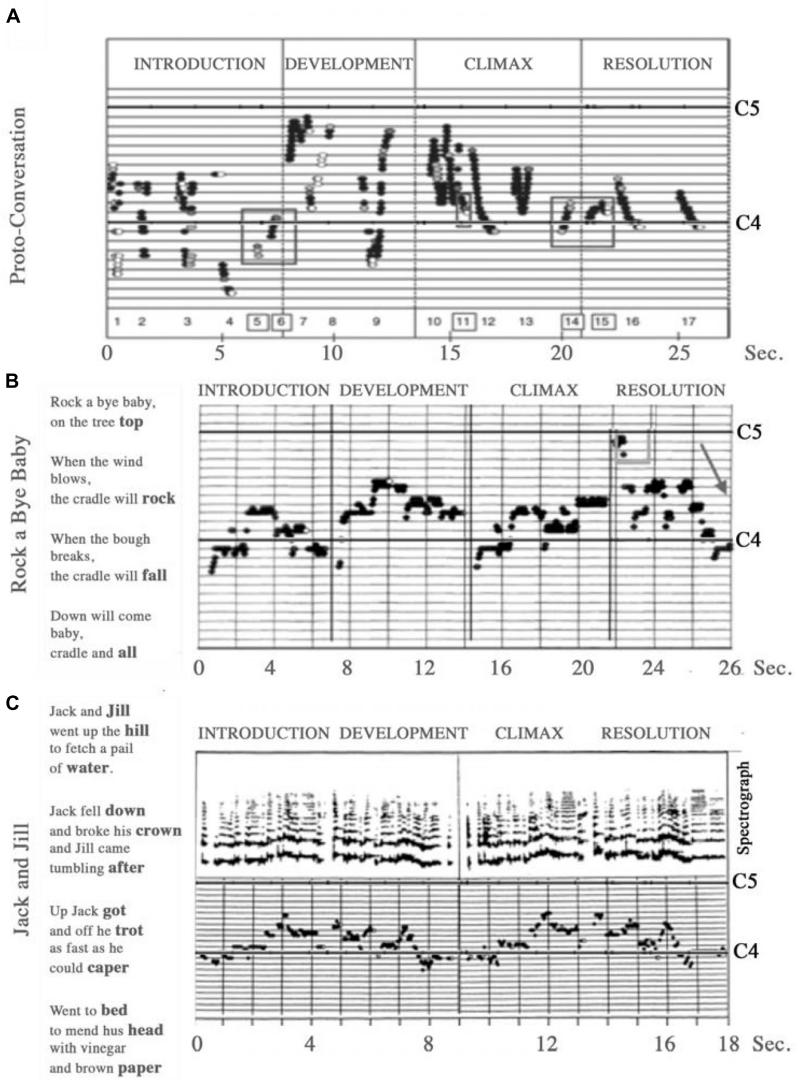



**FIGURE 3 |** Stories of changing pitch in the singing voice. These examples illustrate the narrative phases that infants respond to, which are found in infant directed speech or baby songs in different languages. **(A)** A proto-conversation of a mother with a 6-week-old. The baby shares the development, climax and resolution of the narrative, with vocalizations close to middle C. **(B)** A slow lullaby, of 26 s, has four-line stanzas, with rhyming vowels. **(C)** A more animated song of 18 s shows the same poetic organization with undulations of pitch and vowels rhyming between first and second lines and between final words of each verse. **(A)** Original in Malloch (1999), Figure 5, p. 41. **(B)** From Trevarthen (2015), a portion of Figure 4, p. 137, and Trevarthen (2016), Figure 3, p. 10. Original in Malloch (1999), Figure 10, p. 49. **(C)** Original in Trevarthen (1999), Figure 2, p. 183.

First they looked surprised; then they looked about as if someone had come into the room; and finally they smiled with delight and started performing with the music, inspired by the pulse and melody, joining the music with their different abilities to dance and sing (**Figure 4**).

## COMMUNICATIVE MUSICALITY AND EDUCATION INTO THE CULTURE OF MUSIC

Mastery of a musical culture, and of language, starts with the intuitive vocal interactions between caregiver and infant (Vygotsky, 1966). Our innate communicative musicality is the 'raw material' for cultural forms of music and the rules of grammar and syntax. A child makes stories in sound as an active participant whose pride to belong to the rich musical traditions of society propels them into learning and creating. This is the cultivation of communicative musicality to music, from innate self-expression to cultural practice and a musical identity (MacDonald et al., 2002; MacDonald and Miell, 2004). It is brought to life, as language is, with the enthusiastic support of more experienced companions (Bruner, 1983).

The desire for cultural participation is evident in informal learning in which children's own musical culture grows from the vitality of *The Muse Within* (Bjørkvold, 1992). It is nurtured



**FIGURE 4 |** Katerina Mazokopaki's baby dancers: Upper Left: Georgos, 3.5 months lying on a comfortable bed, responds with a big smile and gestures of hands and feet to accompany the music. Upper Right: Katerina, 9 months, smiles, bounces and extends her arms to 'fly' into action. Lower Left: Panos, 9 months, smiles his greeting, then beats the floor rhythmically with his right hand. Lower Right: Anna, 10 months, standing in her cot, smiles and starts dancing vigorously, swinging her bottom. All four sing with the music (Photos supplied to the CT).

from the music, live and recorded, that the child hears all around and contributes to spontaneously, along with the invention of talking and verse-making with playmates, often accompanied by rhythmic stamping, hopping and jumping (Chukovsky, 1963). Our earliest shared signing of communicative musicality in infancy becomes dialogic 'musical babbling' from around 2 months old (Trevarthen, 1990a). Already at 2 months the infant is learning the cultural gestures and preferences that become the tools through which cultural meaning will be created and exchanged (Moog, 1976; Tafuri and Hawkins, 2008). Lullabies are sung more often in cultures that value quiet infants, playsongs in cultures that value lively enthusiasm in infants (Trehub and Trainor, 1998). A musical 'proto-habitus' is created (Gratier and Trevarthen, 2008).

The infant is born ready to interact and discover her musical culture. Hearing responds to musical sounds from the third trimester of pregnancy (Busnel et al., 1992), and infants can recognize music they heard before birth (Hepper, 1991). They recognize musical contours and rhythmic patterns (Trehub and Hannon, 2006), and 'dance' to music before they are one year old (Zentner et al., 2010). Infant-inclusive singing is preferred, like infant-inclusive speaking[1].

From about 3 months of age in many (probably all) cultures, mothers start to sing baby songs (Trevarthen, 1986) to their infants. In a previous publication the second author summarized their characteristics:

(1) "The beat is clear and precisely modulated; usually in the *andante range* near *moderato (90-100* beats per minute).
(2) Often the beat is marked by clapping, or by regular movements of the baby's body.

[1]We regard the term 'infant directed,' usually used in this context, to be a misnomer that does not acknowledge the co-operative and shared nature of the interaction. We use 'infant-inclusive' to refer to the style of speech and singing that spontaneously occur between caregiver and infant.

(3) The mother sings with a clear melodic tone varying her pitch in a treble range and making rises and falls to make a simple development of the 'emotional narrative'.
(4) Songs commonly have a clear stanza form — usually 4 lines, each of 3 or 4 beats, making a verse (which therefore lasts 8-15 seconds).
(5) Especially in the last line, well controlled musical tricks varying the beat (*rubato*, *sforzando, ritardando, accelerando*) are used to mark a climax and resolution." (Trevarthen, 1989, p. 96).

More recognizable musical forms grow with the spontaneous singing of young children as they play alone or with others (Barrett, 2011), practicing their musical craft. The Norwegian musicologist Jon-Roar Bjørkvold (1992) collected and studied the songs of 4–7-year-olds in three kindergartens in Oslo. He observed how they gave voice to emotion, conveyed information, and established relationships through learning and creating their own children's musical culture. He identified two types of children's singing – 'egocentric' for private pleasure, which, as the child matures, gives way to more social or 'communicative' music making.

As young children mature so they use their voice with a singing kind of expression in progressively more 'symbolic' ways. *Fluid/Amorphous Songs* "evolve in a completely natural way from the infant's babbling as part of its first playful experiments with voice and sound. This type of spontaneous song, with its fanciful glissandi, micro-intervals, and free rhythms, is quite different from what we adults traditionally identify as song." (Bjørkvold, 1992, p.65). *Song Formulas*, such as teasing songs, are symbolic forms for communicating and they flourish after the child begins to play with peers, typically at two or three. Elements of musically more complex *Standard Songs* are picked up from play with adults and hearing them sing, and are adapted to fit what the child is doing. This progressive 'ritualisation' of vocal creativity clarifies the adaptive motives for learning to sing, and how they express increasing narrative imagination for sharing ideas in culturally specific ways (Gratier and Trevarthen, 2008; Eckerdal and Merker, 2009), paralleling the way language is mastered (Chukovsky, 1963).

All through the development of children's singing, repetition and variation, basic tools of any piece of music (for example, see Ockelford, 2017), are primary features as children explore the possibilities of musical form. Repetition and variation between the vocalizations of infant and caregiver feature from the very first shared vocalizations, regulating feelings in social interactions (Malloch, 1999). Later, the growing child will continue to play with how music can convey affect and change their own and others' mood, the four-part structure of Introduction, Development, Climax and Resolution, identified above in the structure of a proto-conversation, becoming the basis of large scale musical works, as well as verbal argument (for example, see the sections of classical rhetoric).

How the child's spontaneous musicality, as it grows in group practice without formal training (MacDonald and Miell, 2004), is received by the surrounding educational culture, is a vital ingredient in the child's emerging 'musical identity.' Musical identity and self-efficacy or mastery of skill in music making inform each other, in reciprocal relationship. A child who sees themselves as a competent musician may attempt to learn a difficult piece of music, and their success at performing this piece will further bolster their sense of competence. And the way a child is welcomed into their musical culture is of vital importance as to whether this child thrives playfully with the musical tools at her disposal, developing her skill in the use of these tools, or shrinks away in disinterest because her own intrinsic musicality is not being heard or valued. If education does its job well, with the child as collaborative artist and thinker (Trevarthen et al., 2018), our rich inner narrative of affective life, generated with our prospective awareness of body movement, is expressed in our social group to create a life-affirming, inclusive culture of shared artful rituals that celebrate the aesthetic grace and moral graciousness of joy in performance (Frank and Trevarthen, 2012; Trevarthen and Bjørkvold, 2016). For example, the InCanto project (Tafuri and Hawkins, 2008) is a wonderful example of infants' and parents' being encouraged to have their expression of music cultivated in such a way that the infant grows into a child who shows greater ability to sing in tune, a greater range of musical expression, and overall more enthusiasm for music participation.

Problems from introducing an emphasis on enforced cultural learning too early are demonstrated by Bjørkvold (1992) who studied the musical games of children in Oslo, Moscow, St Petersburg, and Los Angeles where educational, cultural, social and political practices are very different. In all three countries children showed spontaneous musicality, but in the nations of Russia and the US, where formal training in music was given greater value than it was in Norway, he found reduced spontaneous music making. He insists, "It is critically important for children to master spontaneous singing, for it is part of the common code of child culture that gives them a special key to expression and human growth" (Bjørkvold, 1992, p. 63). A comparable inhibitory effect of conventions of schooling has been recorded on the spontaneous expression of religious feelings and spirituality in the early years (Hay and Nye, 1998). These innate sources of human imagining in collaborative, moral ways give value and meaning to the later cultivation of advanced cultural ideas and skills (Valiente et al., 2012).

The importance of valuing both the child's innate musical creativity and introducing a child into his musical culture so that he may thrive within it and contribute to it can be conceptualized as a balance between two educational necessities – providing a social environment where a child's own skills and abilities are nurtured, and a place where training is provided into the ways of a particular culture (Rogoff, 2003). Both build enthusiasm for cultural participation. This balance has been presented by Bowman (2012) and others as a consideration of two Latin roots for the English word 'education.' One, *educare*, means to train or to mold. Its motivation is the initiation of a person into cultural conventions, without which a person is left unable to live effectively within a particular culture, using its tools to communicate. The other, *educere*, means to lead out, or draw out. Without this more responsive nurturing, the person is left unable to engage with situations and solve problems not yet

imagined. Their ability for creativity is compromised. These very different concepts of what education means are often experienced in schooling as being in tension, with *educare* often winning out, leaving the child with dry knowledge rather than living abilities supported by their own innate skills.

We propose that teachers and students of music at all levels learn how best to do their work by deliberately invoking the rhythms of the student's innate creative vitality while demonstrating cultural conventions that make rich use of this talent (Flohr and Trevarthen, 2008). Infants and toddlers make imaginative musical play in affectionate friendships with parents or peers (Custodero, 2009); primary school children build relationships with the invention of stories in groups with free instrumental play and dance (Fröhlich, 2009); and an advanced music student is assisted to master their instrument through their teacher encouraging their playing to be like a dance representing a narrative, rich in expressive feelings (Rodrigues et al., 2009). In all instances the motives of the learner, and how they may change with development of the body and experiences gained, are of crucial importance (Bannan and Woodward, 2009; Ingold, 2018). As with all education, the success of teaching depends on recognition of how children's 'zest for learning' (Whitehead, 1929; Dewey, 1938) changes with age and the development of body and mind.

We end this section with a quote from Bowman on the broader role of music *in* education. In times where the arts are often considered of marginal importance in education, it talks to the richness of our engagement with music in nurturing all learning experiences:

> "The distinctive educational and developmental potential of music lies, I submit, in dynamic, bodily, and social natures, and distinctly ethical, responsive, and responsible kinds of know-how these afford. Practical knowledge is action embedded knowledge, quite distinct from theoretical knowledge and technical know-how. It is a kind of character-based sense of how best to proceed in situations where best courses of action cannot be determined by previous ones. This ability to discern the right course of action in novel, dynamic situations is precisely the kind of human asset required in today's rapidly changing world. And musical engagements may, under the right circumstances, nurture this capacity in ways unmatched by any other human endeavour." (Bowman, 2012, p. 31).

## COMMUNICATIVE MUSICALITY AND RESILIENCE OF THE HUMAN SPIRIT

As Daniel Stern has written (Stern, 2010), the human body has a rich range of gestural 'forms of vitality' – we move in musical ways. And within each actor there is both 'self-sensing' and 'other-sensing' of the degree of grace, or biological efficiency (Bernstein, 1967) and hopefulness (Trevarthen and Malloch, 2017a) in the gestural narratives of our projects. These qualities of vitality, or well-being, transmitted to others, become the qualities of relationships and social activities – their moral values (Kirschner and Tomasello, 2010; Narvaez, 2014; Trainor and Cirelli, 2015). They convey relational feelings for the degree of consensuality or sharing of expression in moving. Effort to manage the grace and morality of movements can be cultivated to assist well-being of those whose actions are confused or fearful - that is, the making or 'poetry' (from the Greek *poiein*, to make) of their movements may be enhanced to provide responsive and relational care or therapy (Hobson, 1985, ch3; Stern, 2000, p. xiv; Osborne, 2009b; Meares, 2016).

At times our healthy ability for graceful gesturing is met with circumstances that do not allow it to be expressed with its natural healthy vitality. For example, failure to gain a sympathetic appreciation of their musicality can cause an infant to express withdrawal and distress (Murray and Trevarthen, 1985). Instead of joyful pride in sharing play they show sadness and shame (Trevarthen, 2005b). However, an infant's communicative musicality can also be expressive of resilience and determination.

In the example presented in **Figure 5A** we see a consistent rigidity of expression and a lack of self-confident invention on the part of a mother suffering from BPD (borderline personality disorder). She repeats the same up-and-down vocal gesture again and again, with almost no vocal participation on the part of the infant. Where the infant does participate (shown by vocalizations with either a square or circle around them), the infant appears to be setting up the possibility for a dialog – vocalizing exactly on the 'bar-line' (bar 5, shown by a square) and then around the mother's pitch (shown by a circle). Indeed, the infant's vocalizations persuade the mother out of her repetitiveness – the mother momentarily takes notice of her infant and responds to her infant's conversational offering by ceasing her unresponsive repetition and vocalizing once more at the infant's pitch. But the dialog almost immediately breaks down, and the mother returns to her stereotypical, repetitive vocal gesturing.

As well as showing inflexible, 'non-graceful' behavior of a mother who is suffering from BPD, this example also shows the resilience and hope of the infant in the face of her mother's rigid communicative style. This will of our communicative musicality, evident from the earliest of interactions with infants (Papoušek, 1996), is utilized in therapies that employ in-the-moment intersubjective interactions as means for healing (Stern et al., 1998; Stern, 2000, p. xiv). This immediately responsive interaction may be through talking, through music-making, through dance, or touch. The common element is an individual who is highly skilled in attuning to nuances of interpersonal timing and gesture, and who aims to lead back to health another whose personal mindedness of time-in-the-body has become compromised through hardship, suffering, or biological disruption, perhaps leading to a sense of isolation and misunderstanding (Trevarthen and Malloch, 2000). The therapist joins with the person who needs help, leading them back to health and wellbeing through their own therapeutic sense of the 'minute particulars' in that moment of meeting (Hobson, 1985, part 2).

Olga Maratos, a developmental psychologist who pioneered recognition of the ability of young infants to imitate expressions of an attentive and supportive adult (Maratos, 1982) took part in the establishment in Athens of a residential school for children with autism, called *Perivolaki*, meaning 'a little garden.' It is a day-care center with facilities in beautiful surroundings that

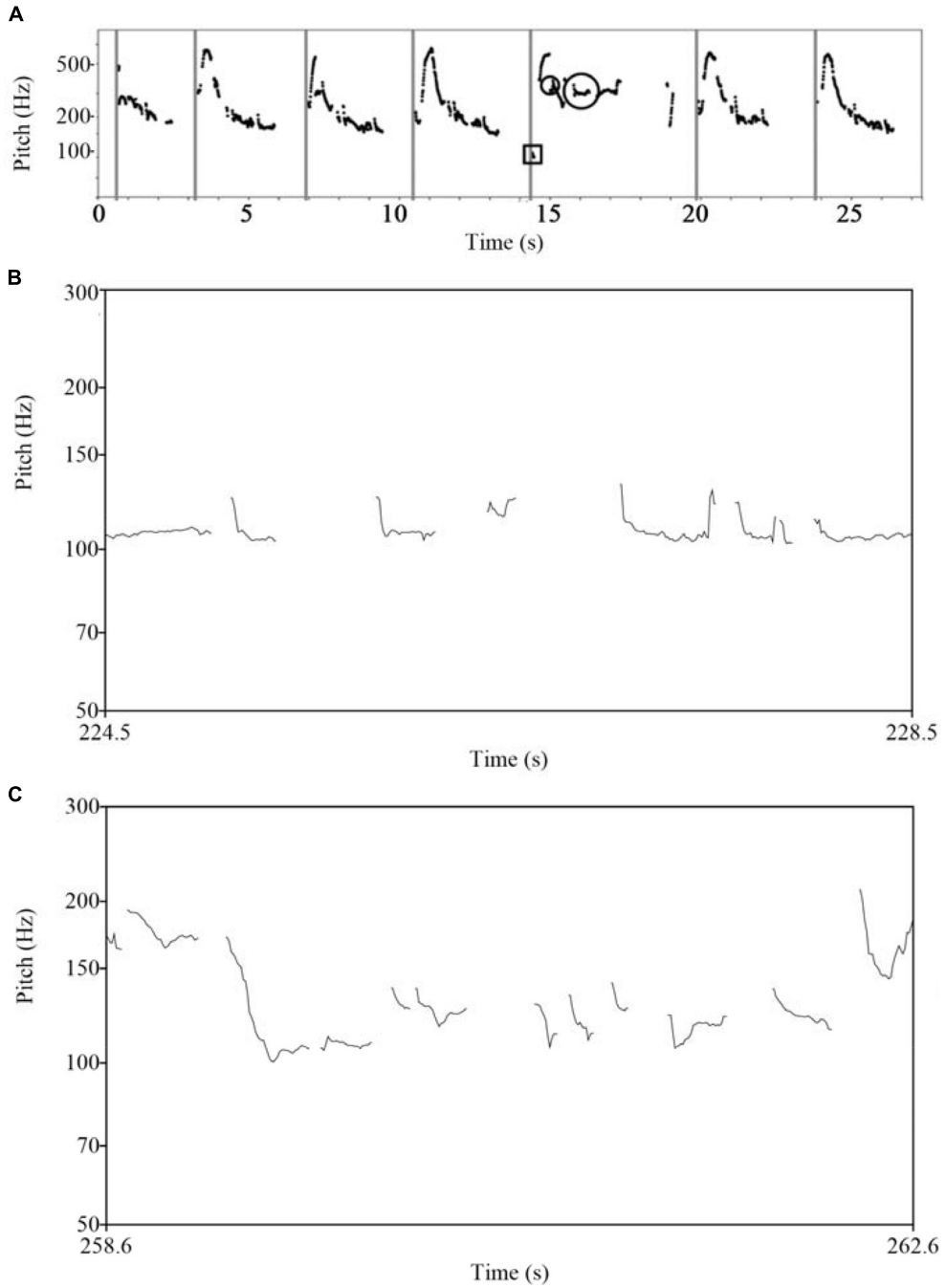



**FIGURE 5 |** Voice modulations that express emotions of relating in a psychological disorder, and in relational therapy. **(A)** A pitch plot of a mother suffering from BPD (borderline personality disorder) speaking to her infant. She repeats the same phrase with monotonous intonation, making a slight reaction to critically timed sounds made by the infant, but cannot establish shared affective engagement. **(B,C)** Changes in the prosody of a client's speaking in communication with a therapist, before and after a therapeutic change, with improvement in self-confidence. **(A)** Original in Gratier and Apter-Danon (2009), Figure 14.5, page 319. Reproduced with permission. **(B,C)** Originals in Malloch (2017), Figures 4.1 and 4.2, p.72. Reproduced with permission.

invite playful out-door activities, games with toys and creative shared occupations such as listening to and making music. They use the psychoanalytic concept of 'transference' of feelings or 'unconscious desires' to encourage sensitive intimate and consistent relations, each child having a trusting relationship with a particular member of the staff. Stability of activities is maintained, with close engagement with the parents, who are seen weekly during the first two years of the child's stay at *Perivolaki* and every fortnight thereafter. The average length of stay is 4–5 years (Maratos, 1998).

Olga explains that staff are trained to observe the children, think about them and discuss their behavior at staff meetings "with a view to understanding whether their autistic behaviour is defensive, refusing interaction and relations because they don't make sense for them or because they are painful, or whether there is a pervasive lack of motivation for relating and communicating. We find both conditions present, at different times, in all our children" (Maratos, 1998, p. 206). This approach, adapting a psychoanalytic treatment which avoids diagnosis of the cause of the disorder, leads to a form of active and intimate 'relational therapy,' which does not rely on verbal formulation of anxieties and lack of trust.

Affect attunement has been defined as qualities of vocal and body gesture that carry meaning in parent–infant communication – it is, "the performance of behaviours that express the quality of feeling of a shared affect state, but without imitating the exact behavioural expression of the inner state" (Stern, 1985, p. 142). This largely unconscious 'recasting' of events is necessary to "shift the focus of attention to what is behind the behaviour, to the quality of feeling that is being shared" (Stern, 1985, p. 142). We say the relationship is now one of 'companionship,' a word from Latin meaning 'to break bread with' and defined here as the wish to be with an other for a mutually beneficial 'inner' purpose, apart from reasons of immediate survival, procreation or material gain. Companionship involves exchanging affect through sharing the quality or virtue of impulses of motivation, which is the original rich meaning of 'sympatheia' in Greek (Trevarthen, 2001).

The therapeutic relationship, even in talking therapies focussed on what the language says about feelings, is underpinned by the manner or graciousness of our gestural exchanges – whether those qualities are carried in vocal prosody or bodily movements (Stern, 2000, p. xiv; Malloch, 2017). It is about the direct and desired sharing of feelings in human vitality. Stern (2010) writes eloquently on the importance of listening for the forms of vitality (the forms of feeling as Hobson, 1985, called them) that are being expressed through prosody – "without the dynamic vitality features of the intention-unfolding process we would not experience a vital human being behind the words that are being said" (Stern, 2010, p.124). He also writes of the use of metaphors as carriers of images of our in-the-moment vitality (also see Hobson, 1985, ch.4). For example, in the work of Stephen, the first author, as a therapist, here is the text of an exchange between him and a client discussing the client's sense that his life has never measured up to his notion of what a successful life should look like:

> Client: My father thought I was stupid. He'd call me 'stupid boy.'
> Therapist: The way you said that – perhaps a sort of exhaustion and emptiness – reminds me of the tide going out in a bay.
> Client: Empty. . . like something's slowly leaving.

Here the focus on the vitality that the client associates with his father's dismissing statement leads to a new experience to explore therapeutically – something is slowly leaving. Further discussion focussed on the experience of what 'slowly leaving' feels like. When the therapist works like this, "we move from an enquiry about intentions, means, and goal states to an enquiry about processes of creation, emerging, and becoming" (Stern, 2010, p.126). We move from distanced recollection or speculation to life in the present moment (Stern, 2004).

The prosody of the client's voice sometimes sums-up the therapeutic change itself. In the example below of Stephen's work a client is talking of an emerging 'new me' in contrast to an 'old me.'[2] The 'old me' was marked with 'a lack of self-respect.' 'I blame myself when things go wrong, I believe I'm not working hard enough.' The voice is drone-like, body hardly moving. **Figure 5B** shows a four-second section of a pitch plot of 'old me' voice.

After describing 'old me' the client's body relaxed, they looked up from the floor, hands lifted from their lap, the volume of their voice increased, its pitch lifted, and they began talking of 'new me.' 'New me is more rational about life. This part says, "Well, I was uncommunicative this morning – that's all right, that's OK. That's just the way I was. Doesn't make me a bad person. Other times I communicate really well!" **Figure 5C** is a four-second pitch plot of 'new me.' The shift in the vitality of the musicality is clear. Stephen *felt* the distinct difference in the vitality of the two "me's" of the client, and continued the session exploring how the new "me" might express itself in the world (see Malloch, 2017, for further discussion of communicative musicality in therapy).

The role of our communicative musicality in supporting our wellbeing lies at the very heart of the practice of using music therapeutically. Music therapy covers a wide range of ways of using musical experiences – stories in sound – to heal and improve people's lives. The Australian Music Therapy Association defines Music Therapy as: "a research-based practice and profession in which music is used to actively support people as they strive to improve their health, functioning and wellbeing." It is the compassionate use of music to engage another emotionally, interpersonally, cognitively, and culturally. "Music is therapeutic because it attunes to the essential efforts that the mind makes to regulate the body, both in its inner neurochemical, hormonal and metabolic processes, and in its purposeful engagements with the objects of the world, and with other people" (Trevarthen and Malloch, 2000). This is particularly so during improvisational music therapy, where the therapist supports the client towards change – greater integration of experience and freedom in communication (Pavlicevic, 2000, ch.6; also see Malloch et al., 2012, on the effectiveness of improvisational music therapy with neonates).

However, the practice of music therapy is more than the therapeutic use of preverbal protomusic, however, important this is. Reflecting our discussion above on music and education, music therapy is also making use of the cultural forms of our musicality, and the power these cultural forms have within our psyche (for example, Donald, 1991; MacDonald et al., 2002; Stern, 2010).

A relationship between our communicative musicality and our culturally made music for the practice of music therapy is proposed by Pavlicevic and Ansdell (2009). They emphasize the peculiarly *musical* relationship established within music therapy practice – that is, that the cultural elaboration of communicative musicality relates to our communal, social lives. Music therapy

[2]Written consent was obtained from the client by the first author.

engages our shared communicative musicality, and welcomes us into the shared cultural, communal experience of musicking, using the tools of a particular cultural type of music – one of many musics in the world (Stige, 2002).

Thus, part of the reason music therapy 'works' is its invitation for cultural collaboration – we exercise what has been called our 'deep social mind' (Whiten, 2000) following particular cultural forms. This wish to learn the forms of culture, our 'conformal motive' (Merker, 2009), comes to life within an environment where we sense our communicative gestures are being valued by another or others through the creation of shared narratives of vitality forms (Stern, 2010). We feel ourselves to be both a companion in our shared narrative of communicative embodied gestures and a companion in a particular shared cultural collaboration. It is within this dual companionship that our deep yearning to belong is met and satisfied, and where healing can occur.

## MUSICAL AFFECTIONS OF THE EMBODIED HUMAN BRAIN

In our opinion brain science has been most insightful into the nature of the self and what makes us human, and how we share the joy and pains of life, when it investigates 'Primary Process' emotional guidance of brain growth for regulation of vitality in body movement and its 'seeking' awareness (Hess, 1954; Solms and Panksepp, 2012). It offers us insight when it investigates how experiences develop by generating expectations of well-being in companionship and by enriching it with cultural meaning (Trevarthen, 1990b). It shows that differences in the rate of development of cognitive processes in left and right hemisphere at different ages are caused by different affections (Thatcher et al., 1987; Chiron et al., 1997), from which arise correlations with musical behaviors and other creative forms of play, as well as Piagetian stages of rational mastery of the body and objects it uses, and development of language.

The neuroscientist Jaak Panksepp, who studied the emotions of mammals who do not cultivate music, but who use patterns of movement including signs with sound to regulate their social lives in ways that anticipate our richer experience of the sounds of our movement using the tools of culture, offers insight into why we are musical beings (Panksepp and Bernatzky, 2002). And in a recent synopsis he agrees that mastery of language in early years depends on the sense of purpose we share in musical ways:

> "Human languages are coaxed into the brain, initially by the melodic intonations of motherese by which emotional communication becomes the vehicle for propositional thought." (Panksepp et al., 2012, p. 11).

Like the brain of any animal, the human brain grows to represent and regulate a body form in movement (Trevarthen, 1980, 2001, 2004). And from even before birth, the self-formation of a personal self in the brain of the fetus is led by manifestations of movement.

> "The first generalized movements occur in week 8 (de Vries et al., 1984), but already in week 5 monoamine transmission pathways grow from the brainstem to animate the primordial cerebral hemispheres. Key components of the Emotional Motor System (hypothalamus, basal ganglia and amygdala) are in place when the neocortex is unformed." (Trevarthen, 2001, p. 26).

After the baby is born and seeks intimate communication of all motives with a parent, the affective system remains as the director of learning and appreciation of what is gained by new awareness.

Psychiatrist and literary scholar Iain McGilchrist in *The Master and His Emissary* (McGilchrist, 2009) has presented a brilliant review of behavioral and brain research, and a clear conception of complementary consciousness in the two cerebral hemispheres. "Music", he writes, "being grounded in the body, communicative of emotion, implicit, is a natural expression of the nature of the right hemisphere" (p. 72).

McGilchrist's research leads him to the political view that we are living in a society that grants too much power to the special refined perceptuo-motor and scientific skills of the left brain, while failing to appreciate how the right brain gives purpose and value to all that we do, thus pointing to the importance of closer cooperation between science and the fundamental values of the humanities. He draws on anthropological information about the universal principles of social understanding at very different levels of technical ability and manufacture, and on the importance in all social groups and cultures of musical performances, which he concludes, from a wide range of evidence, that music evolved before language and contributed to the formulation of its syntax and prosody (see Sachs, 1943).

Music, he says, was not,

> "an irrelevant spin-off from something with more of a competitive cutting edge – namely, language.... rather the reverse. If language evolved later, it looks like it evolved from music.... Rousseau in the eighteenth century, von Humboldt in the nineteenth century and Jespersen in the twentieth, have thought it likely that language developed from music.... That we could use non-verbal means, such as music, to communicate is, in any case, hardly surprising. The shock comes partly from the way we in the West view music: we have lost the sense of the central position that music once occupied in communal life, and still does in most parts of the world today.... We might think of music as an individualistic, even solitary experience, but that is rare in the history of the world." (p. 104).

And he quotes neurologist Oliver Sacks, who said:

> "This primal role of music is to some extent lost today, when we have a special class of composers and performers, and the rest of us are often reduced to passive listening. One has to go to a concert, or a church or a musical festival, to recapture the collective excitement and bonding of music. In such a situation, there seems to be an actual binding of nervous systems, the unification of an audience by a veritable 'neurogamy' (to use a word favoured by early Mesmerists) (Sacks, 2006, p. 2528).

Twenty-six years before McGilchrist published his book, the anthropologist Victor Turner, famous for his book *From Ritual to Theatre*, drew on knowledge of the different functions of the hemispheres to identify play with a collaboration between them, in an article entitled "Play and drama: The horns of a dilemma":

> "Current ideas about differences between the left and right hemispheres of the brain provide a basis for speculating about the nature of play. Play encompasses both the rationality and order of the left hemispheric orientation, and the improvisation and creativity of the right. But play also transcends these oppositions, running rings about them as it encircles the brain's consciousness" (Turner, 1983, p. 217).

Good, beautiful and enjoyable, music is created out of poetic play. Indeed, we *play* music (Trevarthen and Malloch, 2017b).

A sense of time in the mind is the fabric from which movements of all kinds are woven into ambitious projects that value elegance with efficiency. It is a manifestation of the 'biochronology' that is essential to the vitality of all forms of life (Osborne, 2017). In *Rhythms of the Brain*, György Buzsáki (2006) presents a wealth of evidence that the brain functions as a coherent rhythmic system, always in synch., and with a rich array of rhythms that are organized to collaborate.

> "At the physiological level, oscillators do a great service for the brain: they coordinate or 'synchronize' various operations within and across neuronal networks. *Syn* (meaning same) and *chronos* (meaning time) together make sure that everyone is up to the job and no one is left behind, the way the conductor creates temporal order among the large number of instruments in an orchestra" (Buzsáki, 2006, p. viii).

From Panksepp and Trevarthen (2009), p. 114:

> "Music is performed with the measure of expressive movements in time, and with tensions created by combining rhythms (Osborne, 2009a). The 'architecture' and 'narration' of moving in psychological time is also displayed with emotional qualities related to vital functions of the body (Trevarthen, 2009). These psychobiological elements of vitality are charted in three bands or ranges of physical or scientific time: (1) for the felt and imagined 'extended present' (from 10 seconds to years); (2) for the conscious 'psychological present' (Stern, 2004), with its serially ordered steps of motor control coupled to the physiological rhythms of breathing and variations in heart rate (0.3 to 7 seconds); and (3) for 'reflex experiences' and 'just noticeable differences' too fast to be regulated by movements that are prospectively controlled in awareness (5 to 200 milliseconds). (For detail and the sources of this description see Trevarthen, 1999).
>
> The time of musical narrative, which Imberty (2000) calls the macrostructure or 'story-without-words' of music, is related to the times of expressive behavior that form 'protonarrative envelopes' of intuitive vocal and gestural play between infants and their mothers (Stern, 1985, 1995; Malloch, 1999). The period corresponding to a stanza or verse of 20 to 40 seconds may be manifested in the brain, as gamma waves or parasympathetic cycles, which control autonomic functions of the heart and breathing. It continues to be active through sleep to produce fluctuating rates of breathing and heartbeat, as well as electrical activity of the cerebral cortex that might be related to the rehearsal and consolidation of memories in dreaming (Delamont et al., 1999). In wakefulness the narrative cycle is charged and modulated for intersubjective meaning with the 'microtonal' and 'microtemporal' variations of emotion that express urgency and facility in skilled control of moving within the voice of a singer or the playing fingers of an instrumental performer, and in the hearing of a listener (Imberty, 1981, 2000; Gabrielsson and Juslin, 1996; Juslin, 1997, 2001; Kühl, 2007; Osborne, 2009a). Music can assist the synchronization of physiological functions of respiration and heart activity and bring improvement in locomotor activity, and it can improve cognitive and memory processes by brain synchronization."

Rhythmic co-ordination by the Intrinsic Motive Pulse (IMP) of the brain holds body movements together in composition of intentions and experiences (Trevarthen, 1999, 2016). It is the medium for all shared experiences and purposes, and for the convivial vitality of music making.

## CODA: THE PHILOSOPHY OF HUMAN VITALITY

In her review of the role of movement and sense of time in the creation of intelligence, Barbara Goodrich, as a philosopher, traces a history of ideas supporting the view that consciousness is founded on emotions for agency, which we argue are the *sine qua non* for music. In opposition to the "implicit philosophical presuppositions inherited from the canon of Plato, Aristotle, Descartes, and Kant, e.g., that consciousness is self-reflective, passive, and timeless," she proposes a natural science view.

> "Western philosophy, however, also includes what might be described as a counter-tradition—and one that is more compatible with empirical biological science than the usual canon. Heraclitus, Spinoza, Schopenhauer, Nietzsche, and especially the 20th century French philosopher and psychologist, Merleau-Ponty, all anticipated aspects of Llinás's and Buszáki's approaches... sketching out a notion of consciousness emerging from motility, and generating new hypotheses for neurophysiological research." (Goodrich, 2010, p. 331).

We have argued that music comes from this very foundation of consciousness in motivated motility, and we underline the importance of a philosophy that acknowledges the motives and feelings of our life, as well as the intelligence we show in relating to persons, other life forms, and objects in our environment, by recalling the achievements of the philosophers of the Scottish Enlightenment – Hutcheson, Hume, Smith and Reid. In line with Goodrich's "counter tradition," their work anticipates the new understanding of the human BrainMind pioneered by Panksepp and Damasio, which gives primary importance to feelings of vitality in movement, and to emotions that express positive and negative affections in sympathetic communication. This is the science of communicative musicality which underpins the music we create and enjoy.

The Scottish philosophers of the 18th Century, led by Francis Hutcheson, held that relationships and social life depend upon a universal human capacity for "innate sympathy," which generates a conscience, a sense of beauty, a public 'common sense' that values happiness and is disturbed by misery, and a moral sense that perceives virtue or vice in ourselves or others (Hutcheson, 1729, 1755; Hume, 1739–1740).

Smith (1759) in his *Theory of Moral Sentiments* took 'sympathy' to designate any kind of 'moving and feeling

with,' whether motivated positively or negatively, and including posturing and acting in the same expressive way as another's body (cf. the work of Stern on 'affect attunement' quoted earlier), and he also imagined experiences of relating and being sensed, as, for example, interrogating one's conscience.
He said:

> "How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it."
>
> "Sympathy... may..., without much impropriety, be made use of to denote our fellow-feeling with any passion whatever." *Part I – Of the Propriety of Action; Section I – Of the Sense of Propriety, Chapter I – Of Sympathy.*

He examined his conscience to understand being a person in relations.

> "When I endeavour to examine my own conduct, when I endeavour to pass sentence upon it, and either to approve or condemn it, it is evident that, in all such cases, I divide myself, as it were, into two persons; and that I, the examiner and judge, represent a different character from that other I, the person whose conduct is examined into and judged of. The first is the spectator, whose sentiments with regard to my own conduct I endeavour to enter into, by placing myself in his situation, and by considering how it could appear to me, when seen from that particular point of view'. The second is the agent, the person whom I properly call myself, and of whose conduct, under the character of a spectator, I was endeavouring to form some opinion."

This picture of a duplex mind regulated by motives of sympathy anticipates the distinction made by William James in 1892 and by Martin Buber in 1923 between a fundamental "I-Thou" state of awareness and the objective "I-It" relations with the physical word we acquire in communication.

Otteson, joint professor of philosophy and economics, and chair of the Philosophy Department, at Yeshiva University, and adjunct Professor of Economics at New York University, has proposed that Adam Smith's *Theory of Moral Sentiments* (1759) has a more profound message for commerce and industry than *The Wealth of Nations*.

> "Smith's picture thus has a clear anti-Freudian thrust: it denies the hydraulic picture of human emotions according to which emotions build up "pressure" that must be "released." Instead, and more plausibly, it conceives of emotions as things that can be controlled and trained by exercising what Smith calls "self-command." The activity of reciprocal adjustment is then repeated numberless times in every person's lifetime, as it is between and among the people in one's community, resulting in the creation of an unintended and largely unconscious system of standards. These standards then become the rules by which we determine in any given case what kind of behavior is, as Smith calls it, "proper" in a situation and what "improper"—meaning what others can reasonably be expected to enter into." Otteson (2000, November 01).

Smith was a great admirer of the messages of music and wrote about the communication of its poetic massages in his essay *Of the Nature of that Imitation which takes place in what are called the Imitative Arts*, published in 1777 (Smith, 1777/1982).

Beginning with his *A Treatise of Human Nature* (1739), David Hume strove to create a natural science of human psychology in opposition to René Descartes' rationalism. He concluded that desire rather than reason motivates our behavior. Anticipating Merleau-Ponty's phenomenology he also argued against the existence of innate ideas, concluding that we know only what we directly experience. He held that inductive reasoning and causality cannot be justified rationally, rather we follow custom and constant relations between ideas rather than logic. He concluded that we do not have a 'conception of the self,' only sensations of being alive. Following his teacher Hutcheson (1729), he believed that ethics are based on feelings rather than abstract moral principles.

Finally, there is a bold clarity in the work of Thomas Reid, the third great follower of the teachings of Hutcheson, and a vigorous debating companion to David Hume. He wrote *An Inquiry into the Human Mind on the Principles of Common Sense* (Reid, 1764).

Reid founded the Scottish School of Common Sense. For him 'common sense' is based on a direct experience of external reality, experience that becomes internal in language, which is based on an innate capacity pre-dating human consciousness, and acting as an instrument for that consciousness. He distinguished the acoustic element from the meanings which seem to have nothing to do with the sounds as such, a state of language, which he calls 'artificial,' that cannot be the primeval one, which he terms 'natural.' He described the way a child learns language by imitating sounds, becoming aware of them long before he or she understands the meaning in the artificial state of contemporary adult speech. If, says Reid, children were to understand immediately the conceptual content of the words they hear, they would never learn to speak at all. Here Reid distinguishes between natural and artificial signs.

> 'It is by natural signs chiefly that we give force and energy to language; and the less language has of them, it is the less expressive and persuasive.... Artificial signs signify, but they do not express; they speak to the intellect, as algebraic characters may do, but the passions and the affections and the will hear them not: these continue dormant and inactive, till we speak to them in *the language of nature*, to which they are all attention and obedience.' (Reid, 1764, p. 52).

'Language of nature' we equate with our embodied moving consciousness – our communicative musicality. An excess of 'artificial signs,' perhaps aimed at increasing productivity, leads to loneliness and ruthless rationality. However, the cultivation of our communicative musicality, in ourselves and others, through playful music, dance, ritual and sympathetic companionship, makes our communal life of shared work of the body and mind creative in more hopeful ways. It restores our common humanity and our connection with all living things.

## ETHICS STATEMENT

Informed consent was gained for all data presented in this paper.

## AUTHOR CONTRIBUTIONS

SM contributed to all sections of the paper, particularly sections Communicative Musicality and Education into the Culture of Music and Communicative Musicality and Resilience of the Human Spirit. CT contributed to all sections of the paper, particularly sections The Genesis of Music in Infancy – A Short History of Discoveries, Case Studies of Infant Musicality, and Musical Affections of the Embodied Human Brain.

## REFERENCES


Bannan, N., and Woodward, S. (2009). "Spontaneity in the musicality and music learning of children," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 465–494.

Barrett, M. S. (2011). Musical narratives: a study of a young child's identity work in and through music-making. *Psychol. Music* 39, 403–423. doi: 10.1177/0305735610373054

Bateson, M. C. (1971). The interpersonal context of infant vocalization. *Quart. Progr. Rep. Res. Labor. Electron.* 100, 170–176.

Bateson, M. C. (1979). "The epigenesis of conversational interaction: a personal account of research development," in *Before Speech: The Beginning of Human Communication*, ed. M. Bullowa (London: Cambridge University Press), 63–77.

Bateson, P., and Martin, P. (2013). *Play, Playfulness, Creativity and Innovation*, 1st Edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139057691

Bernstein, N. (1966). The active search for information: from reflexes to the model of the future. (In Russian) Republished in English by Feigenberg, I. M., & Meijer, O. G. (1999). The active search for information: from reflexes to the model of the future. *Motor Control* 3, 225–236.

Bernstein, N. (1967). *Coordination and Regulation of Movements*. New York, NY: Pergamon.

Birdwhistell, R. (1970). *Kinesics and Context*. Philadelphia, PA: University of Pennsylvania Press.

Bjørkvold, J.-R. (1992). *The Muse Within: Creativity and Communication, Song and Play From Childhood Through Maturity*. New York, NY: Harper Collins.

Blacking, J. (1995). *Music, Culture, and Experience: Selected Papers of John Blacking (Chicago Studies in Ethnomusicology)*. Chicago, IL: University of Chicago Press.

Bowlby, J. (1958). The nature of the child's tie to his mother. *Int. J. Psychoanal.* 39, 1–23.

Bowman, W. (2012). "Music's place in education," in *The Oxford Handbook of Music Education*, eds G. McPherson and G. Welch (New York, NY: Oxford University Press).

Brazelton, T. B. (1961). Psychophysiologic reactions in the neonate. I: the value of observations of the neonate. *J. Pediatr.* 58, 508–512. doi: 10.1016/S0022-3476(61)80184-4

Brazelton, T. B. (1973). *Neonatal Behavioral Assessment Scale (Clinics in Developmental Medicine, 50.)* London: Heinemann Medical Books.

Brazelton, T. B. (1979). "Evidence of communication during neonatal behavioural assessment," in *Before Speech: The Beginning of Human Communication*, ed. M. Bullowa (London: Cambridge University Press), 79–88.

Brazelton, T. B., and Nugent, J. K. (1995). *The Neonatal Behavioral Assessment Scale*. Cambridge: Mac Keith Press.

Bruner, J. (1966). *Toward a Theory of Instruction*. Cambridge, MA: Belknap Press of Harvard University Press.

Bruner, J. S. (1968). *Processes of Cognitive Growth: Infancy. (Heinz Werner Lectures, 1968)*. Worcester, MA: Clark University Press with Barri Publishers.

Bruner, J. S. (1983). *Child's Talk. Learning to Use Language*. New York, NY: Norton.

Bruner, J. S. (1996). *The Culture of Education*. Cambridge, MA: Harvard University Press.

Bruner, J. S. (2003). *Making Stories: Law, Literature, Life*. Cambridge MA: Harvard University Press.

Buber, M. (1923/1970). *I and Thou (transl. W. Kaufmann)*. Edinburgh: T. and T. Clark.

Bullowa, M., ed. (1979). *Before Speech: The Beginning of Human Communication*. London: Cambridge University Press.

Busnel, M. C., Granier-Deferre, C., and Lecanuet, J. P. (1992). Fetal audition. *Ann. N. Y. Acad. Sci.* 662, 118–134. doi: 10.1111/j.1749-6632.1992.tb22857.x

Buzsáki, G. (2006). *Rhythms of the Brain*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195301069.001.0001

Champagne, R. A. (1990). "The semiotics of lévi-strauss: translation as communication," in *The Semiotic Web, 1989*, eds T. A. Sebeok and J. Umiker-Sebeok (Berlin: Mouton de Gruyter), 61–88.

Chiron, C., Jambaque, I., Nabbout, R., Lounes, R., Syrota, A., and Dulac, O. (1997). The right brain hemisphere is dominant in human infants. *Brain* 120, 1057–1065. doi: 10.1093/brain/120.6.1057

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chukovsky, K. (1963). *From Two to Five. (Trans K. M. Morton)*. Berkley, CA: University of California Press.

Custodero, L. A. (2009). "Intimacy and reciprocity in improvisatory musical performance: pedagogical lessons from adult artists and young children," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 513–530.

Damasio, A. R. (1999). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London: Heinemann.

Damasio, A. R. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. San Diego, CA: Harcourt.

Damasio, A. R. (2010). *Self Comes to Mind: Constructing the Conscious Brain*. New York City, NY: Pantheon.

de Vries, J. I. P., Visser, G. H. A., and Prechtl, H. F. R. (1984). "Fetal motility in the first half of pregnancy," in *Continuity of Neural Functions from Prenatal to Postnatal Life*, ed. H. F. R. Prechtl (Oxford: Blackwell), 46–64.

Delamont, R. S., Julu, P. O. O., and Jamal, G. A. (1999). Periodicity of a noninvasive measure of cardiac vagal tone during non-rapid eye movement sleep in non-sleep-deprived and sleep-deprived normal subjects. *J. Clin. Neurophysiol.* 16, 146–153. doi: 10.1097/00004691-199903000-00007

Dewey, J. (1938). *Experience and Education*. New York, NY: Macmillan Co.

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge, MA: Harvard University Press.

Donaldson, M. (1992). *Human Minds: An Exploration*. London: Allen Lane/Penguin Books.

Eckerdal, P., and Merker, B. (2009). "'Music' and the 'action song' in infant development: an interpretation," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 241–262.

Eibl-Eibesfeldt, I. (1989). *Human Ethology*. New York, NY: Aldine de Gruyter.

Eisenberg, N., and Sulik, M. J. (2012). Is Self–Other overlap the key to understanding empathy? *Emot. Rev.* 4, 34–35.

Ekman, P., and Friesen, W. (1975). *Unmasking the Face: A Guide to Recognizing Emotions From Facial Clues*. Engelwood Cliffs, NJ: Prentice-Hall.

Flohr, J., and Trevarthen, C. (2008), "Music learning in childhood: Early developments of a musical brain and body," in *Neurosciences In Music Pedagogy,* eds W. Gruhn and F. Rauscher (New York, NY: Nova Biomedical Books), 53–100.

Frank, B., and Trevarthen, C. (2012). "Intuitive meaning: supporting impulses for interpersonal life in the sociosphere of human knowledge, practice, and language," in *Moving Ourselves, Moving Others: Motion and Emotion in Intersubjectivity, Consciousness, and Language,* eds A. Foolen, U. Lüdtke, T. P. Racine, and J. Zlatev (Philadelphia: John Benjamins), 261–303. doi: 10.1075/ceb.6.11fra

Freud, S. (1923). *Das Ich und das Es (The Ego and the Id)*. Vienna: Internationaler Psycho-Analytischer Verlag.

Fröhlich, C. (2009). "Vitality in music and dance as basic existential experience: applications in teaching music," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 495–512.

Gabrielsson, A., and Juslin, P. N. (1996). Emotional expression in music performance: between the performer's intention and the listener's experience. *Psychol. Music* 24, 68–91. doi: 10.1177/0305735696241007

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Goodrich, B. G. (2010). We do, therefore we think: time, motility, and consciousness. *Rev. Neurosci.* 21, 331–361. doi: 10.1515/REVNEURO.2010.21.5.331

Gratier, M., and Apter-Danon, G. (2009). "The improvised musicality of belonging: repetition and variation in mother–infant vocal interaction," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 301–327.

Gratier, M., and Trevarthen, C. (2008). Musical narrative and motives for culture in mother-infant vocal interaction. *J. Conscious. Stud.* 15, 122–158.

Hadamard, J. (1945). *The Psychology of Invention in the Mathematical Field*. Princeton, NJ: Princeton University Press.

Hay, D., and Nye, R. (1998). *The Spirit of the Child*. London: Fount (Harper-Collins).

Hepper, P. G. (1991). An examination of fetal learning before and after birth. *Irish J. Psychol.* 12, 95–107. doi: 10.1080/03033910.1991.10557830

Hess, W. R. (1954). *Diencephalon: Autonomic and Extrapyramidal Functions*. Orlando, FL: Grune and Stratton.

Hobson, R. F. (1985). *Forms of Feeling*. London: Tavistock.

Hume, D. (1739–1740). *A Treatise on Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning Into Moral Subjects*. London: The Floating Press.

Hutcheson, F. (1729). *An Inquiry into the Original of Our Ideas of Beauty and Virtue: In Two Treatises Concerning Beauty, Order, Harmony, Design. 2. Concerning Moral Good and Evil*. London: J. and J. Knapton.

Hutcheson, F. (1755). "A system of moral philosophy," in *Three Books*, ed. R. Glasgow (London: A. Millar, and T. Longman. (published posthumously by his son)).

Imberty, M. (1981). *Les Ecriture du Temps*. Paris: Dunod.

Imberty, M. (2000). "The question of innate competencies in musical communication," in *The Origins of Music*, eds N. L. Wallin, B. Merker, and S. Brown (Cambridge, MA: MIT Press), 449–462.

Ingold, T. (2018). *Anthropology And/as Education*. New York, NY: Routledge.

James, W. (1892/1985). *Psychology: The Briefer Course*, ed. G. Allport. Notre Dame, IN: University of Notre Dame Press.

Juslin, P. N. (1997). Can results from studies of perceived expression in musical performances be generalized across response formats? *Psychomusicology* 16, 77–101. doi: 10.1037/h0094065

Juslin, P. N. (2001). "Communicating emotion in music performance: a review and theoretical framework," in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. Sloboda (Oxford: Oxford University Press), 309–337.

Kirschner, S., and Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evol. Hum. Behav.* 31, 354–364. doi: 10.1016/j.evolhumbehav.2010.04.004

Kugiumutzakis, G. (1993). "Intersubjective vocal imitation in early mother-infant interaction," in *New Perspectives in Early Communicative Development*, eds J. Nadel and L. Camaioni (London: Routledge), 23–47.

Kugiumutzakis, G., and Trevarthen, C. (2015). "Neonatal Imitation," in *International Encyclopedia of the Social and Behavioral Sciences*, 2nd Edn, Vol. 16, eds D. James and Wright (Oxford: Elsevier), 481–488. ISBN: 9780080970868. doi: 10.1016/B978-0-08-097086-8.23160-7

Kurth, E. (1991). *Selected Writings (eds L. A. Rothfarb, Trans.)*. Cambridge: Cambridge University Press.

Kukhl, O. (2007). *Musical semantics. European Semiotics: Language, Cognition and Culture, No.* 7. Bern: Peter Lang. doi: 10.3726/978-3-0351-0271-0

Langer, S. K. (1942). *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite, and Art*. Cambridge MA: Harvard University Press.

Lashley, K. S. (1951). "The problem of serial order in behavior," in *Cerebral Mechanisms in Behavior*, ed. L. A. Jeffress (New York, NY: Wiley), 112–136.

Llinás, R. (2001). *I of the Vortex: From Neurons to Self*. Cambridge MA: MIT Press.

MacDonald, R. A. R., Hargreaves, D. J., and Miell, D., eds (2002). *Musical Identities*. Oxford: Oxford University Press.

MacDonald, R. A. R and Miell, D. (2004). "Musical collaboration," in *Learning To Collaborate, Collaborating To Learn: Understanding and Promoting Educationally Productive Collaborative Work*, eds K. Littleton and D. Miell (New York, NY: Nova Science), 133–146.

Malloch, S. (1999). "Mother and Infants and Communicative Musicality," in *Rhythms, Musical Narrative, and the Origins of Human Communication. Musicae Scientiae, Special Issue, 1999-2000*, ed. I. Deliège (Liège: European Society for the Cognitive Sciences of Music), 29–57.

Malloch, S. (2017). "Establishing a therapy of musicality: The embodied narratives of myself with others," in *Companionship: The Rhythm of Relating in Children's Therapies,* eds S. Daniel and C. Trevarthen (London: Jessica Kingsley), 63–81.

Malloch, S., Sharp, D., Campbell, D. M., Campbell, A. M., and Trevarthen, C. (1997). Measuring the human voice: analysing pitch, timing, loudness and voice quality in mother/infant communication. *Proc. Inst. Acoust.* 19, 495–500.

Malloch, S., Shoemark, H., Črnèec, R., Newnham, C., Paul, C., Prior, M., et al. (2012). Music therapy with hospitalized infants—the art and science of communicative musicality. *Infant Ment. Health J.* 33, 386–399. doi: 10.1002/imhj.21346

Malloch, S., and Trevarthen, C., eds (2009a). *Communicative Musicality: Exploring the Basis of Human Companionship*. Oxford: Oxford University Press.

Malloch, S., and Trevarthen, C. (2009b). "Musicality: communicating the vitality and interests of life," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 1–11.

Maratos, O. (1973) *The Origin and Development of Imitation in the First Six Months of Life*. Ph.D. Dissertation, University of Geneva, Geneva.

Maratos, O. (1982). "Trends in the development of imitation in the firsts six months of life," in *Regressions in Mental Development: Basic Phenomena and Theories*, ed. T. G. Bever (Hillsdale, N.J: Erlbaum), 81–101.

Maratos, O. (1998). "Psychoanalysis and the management of pervasive developmental disorders, including autism," in eds *Children With Autism: Diagnosis and Interventions To Meet Their Needs*, 2nd Edn, eds C. Trevarthen, K. J. Aitken, D. Papoudi, J. Z. Robarts (London: Jessica Kingsley), 203–214.

Maturana, H., Mpodozis, J., and Carlos Letelier, J. (1995). Brain, language and the origin of human mental functions. *Biol. Res.* 28, 15–26.

Maturana, H., and Verden-Zoller, G. (2008). *Origin of Humanness in the Biology of Love. Edited by Pille Bullen*. Charlottesville, VA: Imprint Academic.

Maturana, H. R., and Varela, F. J. (1980). *Autopoesis and Cognition*. Dordrecht: D. Reidel. doi: 10.1007/978-94-009-8947-4

Mazokopaki, M., and Kugiumutzakis, G. (2009). "Infant rhythms: expressions of musical companionship," in *Communicative Musicality: Exploring the Basis of Human Companionship, 185-208*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press).

McGilchrist, I. (2009). *The Master and His Emissary: The Divided Brain and the Making of the Western World*. New Haven, CT: Yale University Press. ISBN 0-300-14878-X.

Meares, R. (2005). *The Metaphor of Play: Origin and Breakdown of Personal Being*. New York, NY: Routledge.

Meares, R. (2016). *The Poet's Voice in the Making of Mind*. New York, NY: Routledge. doi: 10.4324/9781315670638

Meltzoff, A. N., and Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science* 198, 75–78. doi: 10.1126/science.198.4312.75

Merker, B. (2009). "Ritual foundations of human uniqueness," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 45–60.

Merleau-Ponty, M. (2012 [1945]). *Phenomenology of Perception*, trans. D. A. Landes. New York: Routledge.

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*. London: Weidenfeld and Nicholson.

Moog, H. (1976). *The Musical Experience of the Pre-School Child*. London: Schott Music.

Murray, L., and Trevarthen, C. (1985). "Emotional regulation of interactions between two-month-olds and their mothers," in *Social Perception in Infants*, eds T. M. Field and N. A. Fox (Norwood, NJ: Ablex), 177–197.

Nagy, E. (2011). The newborn infant: a missing stage in developmental psychology. *Infant Child Dev.* 20, 3–19. doi: 10.1016/j.psyneuen.2014.12.007

Nagy, E., and Molnár, P. (1994). Homo imitans or homo provocans? *Abstract, Int. J. Psychophysiol.* 18:128. doi: 10.1016/0167-8760(94)90394-8

Nagy, E., and Molnár, P. (2004). Homo imitans or homo provocans? Human imprinting model of neonatal imitation. *Infant Behav. Dev.* 27, 54–63. doi: 10.1016/j.infbeh.2003.06.004

Narvaez, D. (2014). *Neurobiology and the Development of Human Morality: Evolution, Culture and Wisdom*. New York: W. W. Norton.

Ockelford, A. (2017). *Comparing Notes: How We Make Sense of Music*. London: Profile Books.

Osborne, N. (2009a). "Towards a chronobiology of musical rhythm," in *Communicative Musicality: Exploring the Basis of Human Companionship,* eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 545–564.

Osborne, N. (2009b). "Music for children in zones of conflict and post-conflict: a psychobiological approach," in *Communicative Musicality: Exploring the Basis of Human Companionship,* eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 331–356.

Osborne, N. (2017). "Love, rhythm and chronobiology," in *Rhythms of Relating in Children's Therapies: Connecting Creatively with Vulnerable Children*, eds S. Daniel and C. Trevarthen (London: Jessica Kingsley), 14–27.

Otteson, J. R. (2000). *Adam Smith: Moral Philosopher*. Available at: https://fee.org/articles/adam-smith-moral-philosopher/

Panksepp, J. (1998). *Affective Neuroscience.: The Foundations of Human and Animal Emotions*. New York, NY: Oxford University Press.

Panksepp, J., Asma, S., Curran, G., Gabriel, R., and Greif, T. (2012). The philosophical implications of affective neuroscience. *J. Consciousn. Stud.* 19, 6–48. doi: 10.1016/j.neubiorev.2010.10.010

Panksepp, J., and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Process.* 60, 133–155. doi: 10.1016/S0376-6357(02)00080-3

Panksepp, J., and Biven, L. (2012). *Archaeology of Mind: Neuroevolutionary Origins of Human Emotions*. New York, NY: Norton.

Panksepp, J., and Trevarthen, C. (2009). "The neuroscience of emotion in music," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 105–146.

Papoušek, H. (1996). "Musicality in infancy research: biological and cultural origins of early musicality," in *Musical Beginnings: Origins and Development of Musical Competence*, eds I. Deliège and J. Sloboda (Oxford: Oxford university press), 37–55. doi: 10.1093/acprof:oso/9780198523321.003.0002

Pavlicevic, M. (1997). *Music Therapy in Context: Music Meaning and Relationship*. *London: Jessica Kingsley*. 4341.

Pavlicevic, M. (1999). *Music Therapy: Intimate Notes*. London: Jessica Kingsley.

Pavlicevic, M. (2000). Improvisation in music therapy: human communication in sound. *J. Music Ther.* 37, 269–285. doi: 10.1093/jmt/37.4.269

Pavlicevic, M., and Ansdell, G. (2009). "Between communicative musicality and collaborative musicing: a perspective from community music therapy," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 357–376.

Piaget, J. (1958). *The Child's Construction of Reality*. London: Routledge and Kegan Paul.

Piaget, J. (1966). *The Psychology of Intelligence*. Totawa, NJ: Littlefield Adams.

Pinker, S. (1997). *How the Mind Works*. New York: W. W. Norton.

Porges, S. W. (2011). *The Polyvagal Theory: Neurophysiological Foundations of Emotions, Attachment, Communication, and Self-regulation*. New York, NY: WW Norton.

Reddy, V. (2008). *How Infants Know Minds*. Cambridge, MA: Harvard University Press.

Reid, T. (1764). *An Inquiry into the Human Mind on the Principles of Common Sense*. London: A. Millar, and Edinburgh: A. Kincaid and J. Bell. (Republished 1997, Edinburgh University Press). doi: 10.1037/11974-000

Rodrigues, H. M., Rodrigues, P. M., and Correia, J. S. (2009). "Communicative Musicality as creative participation: from early childhood to advanced performance," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (Oxford: Oxford University Press), 585–610.

Rogoff, B. (2003). *The Cultural Nature of Human Development*. Oxford: Oxford University Press.

Sachs, C. (1943). *The Rise of Music in the Ancient World, East and West*. New York, NY: W. W. Norton.

Sacks, O. (2006). The power of music. *Brain* 129, 2528–2532. doi: 10.1093/brain/awl234

Sander, L. W. (1964). Adaptive relationships in early mother-child interaction. *J. Am. Acad. Child Psychiatry* 3, 231–264. doi: 10.1016/S0002-7138(09)61920-8

Sander, L. W. (1975). "Infant and caretaking environment: investigation and conceptualization of adaptive behavior in a system of increasing complexity," in *Explorations in Child Psychiatry*, ed. E. James Anthony. (New York, NY: Plenum Press), 129–166.

Sander, L. W. (2008). *Living Systems, Evolving Consciousness, and the Emerging Person*. (eds G. Amadei and I. Bianchi; Psychoanalytic Inquiry), Vol. 26. New York, NY: Analytic Press.

Scheflen, A. E. (1972). *The Stream and Structure of Communicational Behavior*. Bloomington: University of Indiana Press.

Schögler, B., and Trevarthen, C. (2007). "To sing and dance together: from infants to jazz," in *On Being Moved: From Mirror Neurons to Empathy*, ed. S. Bråten (Philadelphia, PA: John Benjamins), 281–302. doi: 10.1075/aicr.68.22sch

Sherrington, C. S. (1955). *Man On His Nature*. London: Penguin Books Ltd.

Small, C. (1998). *Musicking: The Meanings of Performing and Listening.* Hanover, NH: Wesleyan University Press.

Smith, A. (1759). *Theory of Moral Sentiments. Edinburgh (Modern Edition: D. D. Raphael and A.L. Macfie, General Editors, Glasgow Edition)*. Oxford: Clarendon. doi: 10.1093/oseo/instance.00042831

Smith, A. (1777/1982). "Of the nature of that Imitation which takes place in what are called the imitative arts," in *The Glasgow Edition of the Works and Correspondence of Adam Smith, Essays on Philosophical Subjects*, Vol. 3, eds W. P. D. Wightman and J. C. Bryce (Indianapolis, IN: Liberty Fund).

Solms, M., and Panksepp, J. (2012). The "Id" knows more than the "Ego" admits: Neuropsychoanalytic and primal consciousness perspectives on the interface between affective and cognitive neuroscience. *Brain Sci.* 2, 147–175. doi: 10.3390/brainsci2020147

Spitz, R. A. (1945). Hospitalism—an inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Study Child* 1, 53–74. doi: 10.1080/00797308.1945.11823126

Spitz, R. A. (1957). *No and Yes: On the Genesis of Human Communication*. New York, NY: International Universities Press.

Spitz, R. A., and Wolf, K. M. (1946). The smiling response: a contribution to the orthogenesis of social relations. *Genet. Psychol. Monogr.* 34, 57–125.

Stern, D., Sander, L., Hahum, J., Harrison, A., Lyons-Ruth, K., Morgan, A., et al. (1998). Non-interpretive mechanisms in psychoanalytic therapy: the 'something more' than interpretation. *Int. J. Psychoanal.* 79:903.

Stern, D. N. (1971). A micro-analysis of mother-infant interaction: behaviors regulating social contact between a mother and her three-and-a-half-month-old twins. *J. Am. Acad. Child Psychiatry* 10, 501–517. doi: 10.1016/S0002-7138(09)61752-0

Stern, D. N. (1985). *The Interpersonal World of the Infant: A View from Psychoanalysis and Development Psychology*. New York, NY: Basic Books.

Stern, D. N. (1995). *The Motherhood Constellation*. New York: Basic Books.

Stern, D. N. (2000). *The Interpersonal World of the Infant: A View from Psychoanalysis and Development Psychology,* 2nd Edn. New York, NY: Basic Books.

Stern, D. N. (2004). *The Present Moment: In Psychotherapy and Everyday Life*. New York, NY: Norton.

Stern, D. N. (2010). *Forms of Vitality: Exploring Dynamic Experience in Psychology, the Arts, Psychotherapy and Development*. Oxford: Oxford University Press. doi: 10.1093/med:psych/9780199586066.001.0001

Stige, B. (2002). *Culture-Centered Music Therapy,* Gilsum, NH: Barcelona Publishers.

Tafuri, J., and Hawkins, E. (2008). *Infant Musicality: New Research for Educators and Parents. Aldershot, Hants, England*. Burlington, VT: Ashgate.

Thatcher, R. W., Walker, R. A., and Giudice, S. (1987). Human cerebral hemispheres develop at different rates and ages. *Science* 236, 1110–1113. doi: 10.1126/science.3576224

Trainor, L. J., and Cirelli, L. (2015). Rhythm and interpersonal synchrony in early social development. *Ann. N. Y. Acad. Sci.* 1337, 45–52. doi: 10.1111/nyas.12649

Trehub, S. E. (2003). The developmental origins of musicality. *Nat. Neurosci.* 6, 669–673. doi: 10.1038/nn1084

Trehub, S. E., and Hannon, E. E. (2006). Infant music perception: domain-general or domain-specific mechanisms? *Cognition* 100, 73–99.

Trehub, S. E., and Trainor, L. J. (1998). Singing to infants: lullabies and playsongs. *Adv. Infancy Res.* 12, 43–77.

Trevarthen, C. (1974). The psychobiology of speech development. In, E.H. Lenneberg (Ed.), Language and Brain: Developmental Aspects. *Neurosci. Res. Program Bull.* 12, 570-585.

Trevarthen, C. (1977). "Descriptive analyses of infant communication behaviour," in *Studies in Mother-Infant Interaction: The Loch Lomond Symposium*, ed. H. R. Schaffer (London: Academic Press), 227–270.

Trevarthen, C. (1980). "Neurological development and the growth of psychological functions," in *Developmental Psychology and Society*, ed. J. Sants (London: Macmillans), 46-95. doi: 10.1007/978-1-349-16331-1_3

Trevarthen, C. (1984a). "Emotions in infancy: regulators of contacts and relationships with persons," in *Approaches to Emotion*, eds K. Scherer and P. Ekman (Hillsdale, NJ: Erlbaum), 129–157.

Trevarthen, C. (1984b). "How control of movements develops," in *Human Motor Actions: Bernstein Reassessed. Amsterdam*, ed. H. T. A. Whiting (Amsterdam: Elsevier), 223–261. doi: 10.1016/S0166-4115(08)61374-6

Trevarthen, C. (1986). "Development of intersubjective motor control in infants," in *Motor Development in Children: Aspects of Coordination and Control*, eds M. G. Wade and H. T. A. Whiting (Dordrecht: Martinus Nijhof), 209–261.

Trevarthen, C. (1989). "Motives for culture in young children – their natural development through communication," in *The Nature of Culture*, ed. W. Koch (Bochum: Brockmeyer), 80-119.

Trevarthen, C. (1990a). "Signs before speech," in *The Semiotic Web, 1989*, eds T. A. Sebeok and J. Umiker-Sebeok (Berlin: Mouton de Gruyter), 689–755.

Trevarthen, C. (1990b). "Growth and education of the hemispheres," in *Brain Circuits and Functions of the Mind: Essays in Honour of Roger W. Sperry*, ed. C. Trevarthen (New York, NY: Cambridge University Press), 334–363.

Trevarthen, C. (1999). Musicality and the Intrinsic motive pulse: evidence from human psychobiology and infant communication. *Musicae Sci. Fall* 3 (Suppl.), 155–211. doi: 10.1177/10298649000030S109

Trevarthen, C. (2001). Intrinsic motives for companionship in understanding: their origin, development and significance for infant mental health. *Infant Mental Health J.* 22, 95–131. doi: 10.1002/1097-0355(200101/04)22:1<95::AID-IMHJ4>3.0.CO;2-6

Trevarthen, C. (2004). "Brain development," in *Oxford Companion to the Mind*, 2nd Edn, ed. R. L. Gregory (Oxford, NY: Oxford University Press), 116–127.

Trevarthen, C. (2005a). First things first: infants make good use of the sympathetic rhythm of imitation, without reason or language. *J. Child Psychother.* 31, 91–113. doi: 10.1080/00754170500079651

Trevarthen, C. (2005b). "Stepping away from the mirror: pride and shame in adventures of companionship Reflections on the nature and emotional needs of infant intersubjectivity," in *Attachment and Bonding: A New Synthesis. Dahlem Workshop Report 92*, eds C. S. Carter, L. Ahnert, K. E. Grossman, S. B. Hrdy, M. E. Lamb, S. W. Porges, et al. (Cambridge, MA: The MIT Press), 55–84.

Trevarthen, C. (2008). The musical art of infant conversation: narrating in the time of sympathetic experience, without rational interpretation, before words. *Music. Sci.* 12(Suppl.), 15–46. doi: 10.1177/102986490801200 1021

Trevarthen, C. (2009). "Human biochronology: on the source and functions of 'musicality'," in *Music That Works: Contributions of Biology, Neurophysiology, Psychology, Sociology, Medicine and Musicology.* eds R. Haas and V. Brandes (New York, NY: Springer), 221–265.

Trevarthen, C. (2015). Infant semiosis: the psycho-biology of action and shared experience from birth. *Cognit. Dev.* 36, 130–141. doi: 10.1016/j.cogdev.2015.09.008

Trevarthen, C. (2016). The spiritual nature of the infant self: an imaginative actor in relations of affection. *J. Consciousn. Stud.* 23, 258–282.

Trevarthen, C., and Bjørkvold, J.-R. (2016). "Life for learning: how a young child seeks joy with companions in a meaningful world," in *Contexts for Young Child Flourishing: Evolution, Family and Society*, eds D. Narvaez, J. Braungart-Rieker, L. Miller-Graff, L. Gettler, and P. Hastings (New York, NY: Oxford University Press), 28–60. doi: 10.1093/acprof:oso/9780190237790.003.0002

Trevarthen, C., Delafield-Butt, J., and Dunlop, A.-W. (2018). *The Child's Curriculum: Understanding the Natural Talents of Young Children, Allowing Rich Holistic Learning.* Oxford: Oxford University Press.

Trevarthen, C., and Malloch, S. (2000). The dance of wellbeing: defining the musical therapeutic effect. *Norwegian J. Music Ther.* 9, 3–17. doi: 10.1080/08098130009477996

Trevarthen, C., and Malloch, S. (2017a). "The musical self: affections for life in a community of sound," in *The Oxford Handbook of Musical Identities,* 2nd Edn, Chap. 9, eds R. MacDonald, D. Hargreaves, D. Miell (Oxford: Oxford University Press), 155–175. doi: 10.1093/acprof:oso/9780199679485.003.0009

Trevarthen, C., and Malloch, S. (2017b). "Grace in moving and joy in sharing: the intrinsic beauty of communicative musicality from birth," in *Anthropology and Beauty: From Aesthetics to Creativity,* Chap. 8, ed. S. Bunn (London: Taylor & Francis/Routledge), 115–130.

Turner, V. (1982). *From Ritual to Theatre: The Human Seriousness of Play*. New York, NY: Performing Arts Journal Publications.

Turner, V. (1983). "Play and drama: the horns of a dilemma," in *The World of Play. Proceedings of the 7th Annual Meeting of the Association of the Anthropological Study of Play*, ed. F. E. Manning (West Point, NY: Leisure Press), 217–224.

Valiente, C., Swanson, J., and Eisenberg, N. (2012). Linking students' emotions and academic achievement: when and why emotions matter. *Child Dev. Perspect.* 6, 129–135. doi: 10.1111/j.1750-8606.2011.00192.x

van Rees, S., and de Leeuw, R. (1993). Born too early: the kangaroo method with premature babies, video by stichting lichaamstaal. *Scheyvenhofweg* 12:6093.

Vygotsky, L. S. (1966). "Genesis of the higher mental functions. Abridged translation," in *Learning to Think*, eds P. H. Light, S. Sheldon, and M. Woodhead (London: Routledge and Open University Press), 32–41.

Whitehead, A. N. (1929). *The Aims of Education and Other Essays*. New York, NY: Macmillan.

Whiten, A. (2000). "The evolution of deep social mind in humans," in *The Descent of Mind*, eds M. Corballis and S. Lea (Oxford: Oxford University Press).

Zeedyk, M. S. (2006). From intersubjectivity to subjectivity: the transformative roles of emotional intimacy and imitation. *Infant Child Dev.* 15, 321–344. doi: 10.1002/icd.457

Zentner, M., Eerola, T., and Purves, D. (2010). Rhythmic engagement with music in infancy. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5768–5773. doi: 10.1073/pnas.1000121107

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2018 Malloch and Trevarthen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 18 February 2020
doi: 10.3389/fpsyg.2019.03051

# The Overlooked Tradition of "Personal Music" and Its Place in the Evolution of Music

*Aleksey Nikolsky*[1]*, *Eduard Alekseyev*[2,3], *Ivan Alekseev*[4,5] and *Varvara Dyakonova*[6]

[1] Braavo Enterprises, Los Angeles, CA, United States, [2] Independent Researcher, Boston, MA, United States, [3] The State Institute for Art Studies of the Ministry of Culture of the Russian Federation, Moscow, Russia, [4] Experimental Laboratory of the North-Eastern Federal University, Yakutsk, Russia, [5] International Jaw Harp Music Center, Yakutsk, Russia, [6] Department of Art Studies, Arctic State Institute of Arts and Culture, Yakutsk, Russia

This is an attempt to describe and explain so-called timbre-based music as a special system of musicking, communication, and psychological and social usage, which along with its corresponding beliefs constitutes a viable alternative to "frequency-based" music. Unfortunately, the current scientific research into music has been skewed almost entirely in favor of the frequency-based music prevalent in the West. Subsequently, whenever samples of timbre-based music attract the attention of Western researchers, these are usually interpreted as "defective" implementations of frequency-based music. The presence of discrete pitch is often regarded as the structural criterion that distinguishes music from non-music. We would like to present evidence to the contrary—in support of the existence of indigenous music systems based on the discretization and patterning of aspects of timbre, rather than pitch. This evidence comes mainly from extensive ethnographic research systematically conducted in eastern European and Asian parts of Russia from the 1890s. It involved the efforts of thousands of specialists and was coordinated by dozens of research institutions, and it has included not just ethnomusicology but linguistics, philology, organology, archaeology, anthropology, geography, and religious, and social studies. Much of the data has not been translated into Western languages. Although some Soviet-era publications were tainted by Marxist ideology, many researchers strove to provide accurate information (despite at times having been prosecuted for their work), and post-1990 research undertook a substantial revision of ideologically compromised concepts. Timbre-based tonal organization (TO) differs from that based on frequency in its personal orientation: musicking here occurs primarily for oneself and/or for close relatives/friends. Collective music-making is rare and exceptional. The foundation of timbre-based music seems to have vocal roots and rests on "personal song"—a system of personal identification through individualized patterns of rhythm, timbre, and pitch contour, utilized like a "human voice"—whose sound enables the recognition of a particular individual. The instrumental counterpart of the personalized singing tradition is the jaw harp tradition. The jaw harp is the principal musical instrument

OPEN ACCESS

**Edited by:**
*Margarita L. Mazo, The Ohio State University, United States*

**Reviewed by:**
*Jaan Ross, Estonian Academy of Music and Theatre, Estonia*
*Theodore Levin, Dartmouth College, United States*

***Correspondence:**
*Aleksey Nikolsky*
*aleksey@braavo.org*

**Specialty section:**
*This article was submitted to Cognition, a section of the journal Frontiers in Psychology*

**Received:** *28 June 2019*
**Accepted:** *24 December 2019*
**Published:** *18 February 2020*

**Citation:**
*Nikolsky A, Alekseyev E, Alekseev I and Dyakonova V (2020) The Overlooked Tradition of "Personal Music" and Its Place in the Evolution of Music. Front. Psychol. 10:3051. doi: 10.3389/fpsyg.2019.03051*

for at least 21 ethnicities in Russia, who occupy over half the territory of the country. The evolution of its TO forms the backbone for the development of timbre-based music art. Here, we provide the acoustic, socio-cultural, geographic, and chronological overview of timbre-based music.

**Keywords: timbre-based music, personal song, Jaw harp (aka Jew's harp), musicality, arctic hysteria, music evolution, thematicism, musical texture**

In the past few decades, a lively discussion on matters concerning the origin and evolution of music has finally begun to move toward a consensus among specialists (Cross and Morley, 2009): the biological importance of music is being seen in its capacity to foster and sustain social interactions within a group, to the mutual benefit of its members. Here, music stands as an important counterpart to language—another biological marker of *Homo sapiens*—specializing in managing the emotional aspect of human interaction. Without diminishing the importance of this perspective, we wish to cast light as well on the *personal* function of music—its capacity to organize and sustain the psychological identity of an individual. This function must have factored in the evolution of music since at least the Upper Paleolithic.

The need for the update grows as more scholars lean toward regarding music as an exclusively collective phenomenon. Thus, Lewis (2013) concludes: "In most parts of the world, and *for most of human history*, music exists *only* because of the *social relations that enable its performance*." Levinson (2013) infers that in the evolution of music its structural complexity "may have its origins in *joint action* rather than in *abstract representations* or *solitary mentation*"[1]. We will argue the contrary—that at least in the traditional lifestyle of numerous native Siberian ethnoses, music serves *primarily* as means of solitary mentation and abstract representation of reality. There is no reason to believe that this manner of "musicking" (Small, 1998) is a recent development, and there are valid reasons to prototype prehistoric North Eurasian music upon this type of music.

## DICHOTOMY BETWEEN TIMBRE-BASED AND FREQUENCY-BASED MUSIC

Almost everything that we know about the perception of music and what has served as a scientifically established foundation for modern views on the origin of music comes from a musical tradition based on frequency discrimination of musical sounds. It is this tradition that currently prevails in the world. Its prevalence probably started with the rise of Bronze Age urban civilizations, whose palace and temple music traditions relied on math-based theory (Nikolsky, 2016). Rationally defined pitches have made the corresponding music practices rely on the frequency aspect. Civilizations that cultivated frequency-based music imposed their influence on the music cultures of neighboring peoples. On a global scale, this must have resulted in a steady decline of the alternative form of TO that was based on timbre. This process is evident in the music cultures of many native Siberian peoples, e.g., the Nganasan, whose timbre-based tradition has been recently overtaken by the Russian frequency-based tradition (Bicheool, 2009). Schneider (2001) qualifies such development as "pitch reductionism"—the replacement of timbre-based tuning standards by frequency-based standards, most obvious in the indigenous gong/bell and xylophone music. Timbral vocal music is no less vulnerable to pitch reductionism.

**Examples-1/2**[2]

Ex.1. Timbre-based vocal music: throat-singing song, "Seagull," in archaic style, by Anna Ankhani, a 70-year-old Koryak woman from a remote reindeer-breeding settlement, Khailino, in Northern Kamchatka (reachable only by aerial transportation). Singing involves throat-rasping and double phonation that allows the singer to produce sounds an octave below her speaking voice range (http://chirb.it/n5JNvk).

Ex.2. Modern style Koryak "Festive song" by Maria Appolon, a 48-year-old woman from a little town, Ossora, a seaport at Bering Sea. Maria is a legislator in the local district Council and a member of the folk ensemble, "Agya," that often performs at international festivals, special events, and for tourists. Her singing exhibits traits of "frequency-based music": no timbral effects, clear discrete pitches that are interconnected without gliding that is so typical for traditional singing in Kamchatka, hexatonic mode, consisting of 2 motifs (a descending trichord and an ascending tetrachord) which retain amazing precision in intonation, without noticeable fluctuations, and strict formulaic structure (http://chirb.it/FK7A8A).

In the literature in English, the distinction between "pitch-centered" and "timbre-centered" musics was drawn by Levin and Suzukei (2006, p. 45–72). Fales (2002) discussed the downsides of "pitch-centrism" and "timbre-deafness" in approaching timbre-based music. In the Russian literature, the dialectics of timbre and pitch was acknowledged much earlier. The pioneer of ear-training and music psychology in Russia, Maykapar (1900), noticed that the musical ear developed from a timbre- to a frequency-reference frame, and proposed to begin musical education with timbral exercises. Since then, ear-training has become an obligatory part of music education in Russia, and has attracted many researchers (Blium, 1977). Adoption of a state-controlled, unified system of early music education

**Abbreviations:** PS, personal song; JH, jaw harp; AH, arctic hysteria; TO, tonal organization; FF, fundamental frequency; f1, first harmonic; F1, first formant; dBFS, decibels relative to full scale; BP, before present.

[1]The emphasis in both quotations is added to the original.

[2]In case if the internet links provided for the audio examples end up being broken, the archive with all of the audio examples (named "Audio for the article") is available in the **Supplementary Material**.

by the Bolsheviks created an exceptionally diversified cross-cultural pool of subjects for the investigation of music skills acquisition (Arakelova, 2012). Based on such data, Teplov (1947)[3] formulated his theory of interference between "timbral" (speech-like) and "pitch" (music-like) hearing—the latter typically replacing the former by the beginning of a child's primary education, provided there was cultural exposure to frequency-based music[4].

Teplov's theory was substantiated by Leontyev (2009, p. 115–36), Vygotsky's pupil, in a series of experiments that disclosed a "timbre-centered" estimation of pitch in adults who lacked music schooling[5]. To demonstrate that pitch-ear is a "functional organ"[6]—absent at the moment of birth and forming as a consequence of cognitive development—Leontyev built a vibrator machine and successfully taught subjects to discriminate pitch exclusively by touch. Another Vygotskian, Zaporozhets, headed the research on the emergence of frequency-oriented hearing in preschool children (Endovitskaya, 1959, 1964; Lisina, 1966; Mukhina and Lisina, 1966; Repina, 1966a,b; Zaporozhets, 2003)[7]. His group concluded that the genesis of frequency discrimination reflected a fundamental re-organization of motoric, visual, and spatial orientations throughout childhood—embodying a form of abstraction "of a number of perceptual activities, directed at the exploration of objects and phenomena of reality, identification and fixation of their perceptual attributes and interrelations" (Zaporozhets, 1985, p. 1:18).

The frequency-based "musical ear" corresponds to the urban Western environment, where systemically organized straight parallel lines and right angles are widespread and dominant. But these do not exist in the steppe or tundra (Nikolsky, 2016, Appendix-7 in **Supplementary Material**). Life in the tundra promotes music systems based on *indefinite (ekmelic or khasmatonal) pitch* (Alekseyev, 1986) and *fine timbral distinctions*, corresponding to life in an open terrain that lacks landmarks. There, orientation occurs via *qualitative* (non-quantitative!) evaluations of wind, light, snow, distance, etc., rather than the incremental and reversible pathfinding in the urban landscape. Most likely, timbre-centrism originates from life in such natural sound-stages where "definite" rhythm and frequency are mostly bound to human activities, and are overall less important than timbre-differentiated environmental sounds (Fales, 2002). The opposition of indefinite/definite pitch systems in Russian musicology finds its match in the opposition of "smooth" and "striated" pitch in Western ethnomusicology (During and During, 2015).

Timbre-centeredness distinguishes the vocalizations of newborns: crying, screaming, rasping, grunting, whining, sobbing, whimpering, etc. (Loewy, 1995). Even prenatally, fetuses recognize parental voices (Lee and Kisilevsky, 2014), which involves at least some timbral discrimination. Newborns distinguish different musical instruments by timbre (McAdams and Bertoncini, 1997). Timbre discrimination is so crucial for a newborn's life that it appears innate (Simons, 1986, p. 43). Perhaps the ontogenetic progression from timbre to frequency (Teplov, 1947) takes the same course as cultural evolution, and "indefinite-pitch" music systems precede "definite-pitch" systems (Alekseyev, 1986, p. 14–15)—following the general paradigm suggested by Foster (1994). Unfortunately, published research on the acquisition of musical skills overwhelmingly focuses on the experience of children in Western societies[8]. Garfias (1990) seems to share Alekseyev's conviction that in indigenous music cultures infants start their musical development from speech-like timbral hearing. Tuvan children learn early to vocally imitate typical environmental sounds with amazing precision, adopting learned timbral distinctions for the creation of their own music (Levin and Suzukei, 2006, p. 85–7)—very much like Western children model their vocal improvisations upon commonly heard tunes (Bjørkvold, 1992). Although Tuvans also use frequency-based music, it remains secondary in importance for them.

In Western classical music, the opposite pertains: it is timbre that is secondary (Scruton, 1997, p. 77–8). Even its standard definition—the "attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar" (ASA, 1951)—is culturally skewed toward prioritizing frequency. Western "frequency-centrism" is most evident in the tendency to ascribe pitch values to clearly

[3]Recently, considerable material has emerged, which well fits Teplov's theory. Timbre and pitch perception seem to engage the same brain resources, and therefore require selective attention (Melara and Marks, 1990a). Analytic extraction of pitch affects extraction of timbre properties, and vice versa (Melara and Marks, 1990b). Timbre affects judgements of simple pitch discrimination (Singh and Hirsh, 1992). Melodic motion can be compromised by timbral shift so that non-musicians could perceive a melodic tritone as larger in size than a perfect 5th (Russo and Thompson, 2005). Timbre can interfere with perception of melodic direction (Allen and Oxenham, 2014). Mistakes in melodic direction due to changes in timbre and loudness are especially pronounced in children as opposed to adults (Williams, 1990).

[4]The detrimental influence of usage of Western frequency-based music on timbral perception has been experimentally demonstrated in a number of studies. A tonal reference can reduce timbre's interference with pitch extraction (Warrier and Zatorre, 2002). This interference is much greater for Western non-musicians (less experienced in pitch analysis) than for musically trained listeners (Pitt, 1994). On the other hand, simplification and uniformity of spectral content seems to enhance frequency-interval discrimination in Western music (Zarate et al., 2013).

[5]Recently, this claim was substantiated by the evidence for two alternative pitch encoding mechanisms based on frequency resolution or non-resolution of harmonics of the fundamental tone (Grimault et al., 2002). This difference seems to lie at the heart of the distinction between "timbre-centered" and "frequency-centered" listening modes.

[6]Leontyev borrowed the concept of "functional organ" from Aleksey Ukhtomsky (1978) who held that "organ" did not necessarily have to constitute a morphologically distinct anatomical part, but may have constituted an organization of a specific function with *dynamic* rather than static attributes.

[7]Some support for Leontyev's and Zaporozhets' conclusion of post-natal formation of pitch discrimination skills comes from more recent Western research. Thus, McAdams and Bertoncini (1997) found that at birth infants can discriminate between rising and falling melodic contours only when they are presented with *timbrally and spatially contiguous sounds*. Schwarzer (1997) found that 5–7 year old children generally favor tracking loudness and timbre in analytic listening tasks over tracking melodic contour—unlike adult listeners.

[8]At present, cross-cultural investigation of general patterns of acquisition of music skills is in its virginal state. Only a handful of publications in English discuss the issue of musical development in non-Western societies (Blacking, 1967; Garfias, 1990; Fernald, 1992; Papoušek, 1996; Minks, 2002; Stige, 2002; Nettl, 2005).

non-musical sounds (e.g., car brakes)—including sounds of non-Western timbral music, such as Inuit (Walker, 1997, p. 323). Unsurprisingly, Westerners' attempts to reproduce indigenous timbral music introduce distortions and are rejected by native users (Ojamaa, 2005).

The reverse bias affects the use of frequency-centered music by timbre-centered musicians. Soviet researchers of the collectivization/*likbez* era discovered that children (Beliayeva-Ekzempliarskaya, 1925) and adolescents (Antoshina, 1939) who lacked exposure to classical music could not detect a harmonic mismatch when a well-known melody was performed against accompaniment in a wrong key. Transference of popular tunes across different music systems usually involves systemic pitch conversion. Thus, the Russian-Ukrainian diatonic heptatonic song "Provody" becomes anhemitonic pentatonic in Buryat and whole-tone tetratonic in Yakut reproductions, severely distorting the original intervallic structure (Alekseyev, 1986, p. 148–55). Both, Buryats and Yakuts, cultivate timbral music. It might be appropriate here to speak of the intuitive substitution of "pitch classes" of the original foreign song (which was composed circa 1918 in Ukraine within the framework of Western tonality) by the "timbral classes" of traditional Buryat and Yakut music which sounded the closest to the original. The "strange" and complex sounding tonality of the foreign song, which for some reason attracted the attention of the indigenous musicians, was substituted by the TO that was habitually "normal" and easy for them.

Timbre-oriented musical cultures seem to base their musical modes on a set of pitch levels that are joined together according to some common trait(s) of timbral coloration and/or sound production technique(s). Thereby a number of such related "timbre-classes" become united into a "timbral mode." This is in contrast to a specific intervallic distance in pitch that distinguishes one "degree" of a musical mode and/or key from another "degree" in frequency-oriented music cultures. Inevitable conversion from pitch- to timbre-based frame of reference must be responsible for the intervallic distortions, discovered by Alekseyev in a Russian-Ukranian tonal song upon its appropriation by the indigenous Buryat and Yakut musicians.

## PERSONAL NATURE OF TIMBRE-BASED MUSIC

Often overlooked is the fundamental unsuitability of timbre-centered music for collective use—in sharp contrast to frequency-centered music. Even before the crystallization of Western tonality, musicians realized that the more instruments that played the same harmony, the more euphonic it appeared (Mersenne, 1957, p. 270). However, the more speakers who collectively pronounce the same sentence, the less articulate it sounds. People sing together, but take turns speaking (Brown, 2007). Fusion of sounds dominates production and perception of music (Huron, 2001) because pitches readily conjoin in unisons, "double-notes" and chords. Timbre-centered music, however, resembles speech in its soliloquacity: for a native listener, simultaneous throat-singing by a dozen singers would make the song unintelligible.

Unlike the simple dimensionality of height for pitch, the multi-dimensionality of timbre considerably complicates its categorization (Krumhansl, 1989). Timbre-based music also misses the systemic rationality of frequency-based music systems. The interrelation of pitches cross-refers any given pitch value, facilitating the interpretation of pitch by making it "rational"—whereas timbre stands as "irrational," lacking anything akin to intervallic ratios (Balzano, 1986). To add to the confusion, the temporal dimensions of timbre interact with its spectral dimensions (Caclin et al., 2007), disallowing multiple timbres to synchronize perfectly. Different timbres, dubbed in unison, either blend into a new timbre, intensifying one of the constituents, or repel each other (Sandell, 1995). Blending depends primarily on the onset of synchrony, similarity of attacks, and/or spectral centroids—traditionally studied within the field of orchestration (Tardieu and McAdams, 2012). Orchestration is limited to frequency-centered music, where musical instruments are perfected to produce clear pitch. Even many popular instruments exhibit registrally fixed formants, and can therefore be considered "pitch-generalized"—which greatly influences their timbral blend (Lembke and McAdams, 2015). Unsurprisingly, unison blending is greater than non-unison blending, inverse to the identifiability of constituent timbres (Kendall and Carterette, 1993). The fusion of multiple "pitch-generalized" instruments generates orchestral *tutti*—an undistinguishable mass of timbres that nevertheless remains "pitch-clear."

*There is no tutti in indigenous timbre-centered music*[9]. Such music cultivates very different instruments (whip, buzzer, cane, flask), distinguished by a characteristic timbre and indefinite pitch (Mazepus and Galitskaya, 1997). In Russian musicology, these are qualified as "phono-instruments"[10]—sound-producing tools manufactured for some common application other than music-making (Yesipova, 2008). The most important of them is the Jaw Harp (JH)[11]. In the contiguous area from the Urals to the Okhotsk Sea, all indigenous ethnicities have their JH traditions (Emsheimer, 1986). A century ago, every ethnicity of Asiatic Russia possessed the JH (Jochelson, 1928, p. 217). The JH can be qualified as the archetypical instrument of Siberia (Sheikin,

[9]This generalization applies only to the indigenous folk traditions of timbre-based music – and not to the styles of Western classical music that are based on sonoristic effects rather than conventional "melodies" or "chords," such as "sonorism" (e.g., Krzysztof Penderecki) or "spectralism" (e.g., Tristan Murail). Such styles generally tend to follow an "experimental" approach forged by Western modernistic composers through "inventing" their own original means of tonal organization rather than implementing already existing folk timbre-based music systems that have been "naturally" formed over a long period of time within an indigenous cultural tradition.

[10]The term "phono-instrument" was introduced by Sheikin (1996) to supplement the Sachs/Hornbostel structural classification of musical instruments by the classification of functionality of their use, where "phono-instruments" represent the archaic forms of musicking—preceding the invention of instruments designed specifically to generate a particular type of music (Sheikin, 2002).

[11]This lamellophone musical instrument is often called "Jew's harp," which is somewhat confusing, since it is neither connected to Jews, nor to Judaism (Wright, 2015, XIV). A more appropriate name seems to be "jaw harp" (Crane, 1982), as suggested earlier by Sachs (1940, p. 58)—for the characteristic manner of its sound production. Player plucks the lamella while holding the instrument close to his/her mouth, used as a resonance chamber for amplifying the desired partials of the vibrating lamella (Ledang, 1972; Adkins, 1974; Trias, 2010).

1996). It and other "mouth-resonating instruments" (mirlitons, bows, "singing-pipes") *specialize in making "music for oneself"* (Sheikin, 2002, p. 116–66). The JH's solitary use prevails over the Volga Plateau (Zagretdinov, 1997), Tuva/Altai (Suzukei, 1989), Afghanistan (Koskoff, 2008, p. 1062), Sakhalin (Mamcheva, 2005), Taiwan (Blench, 2004), Indochina (Sam, 2008), Indonesia (Matusky, 2008), and New Guinea (Pugh-Kitingan, 1977).

Such a wide geographic distribution indicates the inter-cultural—*acoustic*—reason for JH's personal application: except for the recently invented JH models[12], traditional instruments are barely audible beyond a few meters. Even in social settings, JH remains **private**. Such are JH's romantic responsorial duets between a young man and a young woman, known, other than in Russia, in Southern China (Picken, 1957, p. 154), Tibet (Arcones, 2013, p. 216), Hainan (Hsu, 2001), Taiwan (McGovern, 1922), Laos (Simana and Preisig, 2003), Vietnam (Rault and Brenton, 2000, p. 83), Indonesia (Kartomi, 2012, p. 160), New Guinea (Pugh-Kitingan, 1977), and Western Europe (Kolltveit, 2006, p. 111).

**Example-3**

"Serenading" *khomus* (Yakut JH): a romantic duet in traditional style, performed by Erkin Alekseyev and Tokuiaana Nikolayeva. Music proceeds in a responsorial setting, but ends with simultaneous playing that reflects the union of feelings (http://chirb.it/rH6bFD).

Germane in this context is that "serenading" duets involve taking turns, like interlocution: one player reproduces the sonic material suggested by another player (Haid, 1999).

The JH is also a child's favorite plaything—another cross-cultural sphere of private use—in Yakutia (Dyakonova, 2017), Ural (Aleksandrova, 2017), Sakhalin (Mamcheva, 2012, p. 197), Uzbekistan (Beliayev, 1933), Kyrgyzstan (Vinogradov, 1958, p. 180), Afghanistan (Slobin, 1976, p. 53), Mongolia (Pegg, 2001), Japan (Ishi, 1916), southeastern China (Picken, 1957), Polynesia (McLean, 1999), Indonesia (McPhee, 1955), and Western Europe (Kolltveit, 2006, p. 109). Toying with a JH is typical during the long hours of herding (Shchurov, 1995), a task entrusted to children in nomadic Asiatic societies (Stépanoff et al., 2017). In Altai, the JH is regarded as a herding instrument (Dorina, 2004). Children start working in pastures at ages 5–6 (Yekeyeva, 2011). In Yakutia, mothers teach their children to make JHs from tree splinters (Tchakhov, 2012). Nivkh 5-year-olds learn to play and make bamboo and grass JHs, guided by parents, grandparents, or older siblings[13].

Self-manufacturing is an important marker of timbral instruments; personal use accompanies personal manufacturing (Dorina, 2004)[14]. Even metallic JHs are often self-made, such as by flattening brass rifle cartridge cases (Mamcheva, 2012, p. 50). Forging of metallic JHs by a metalsmith is a relatively recent historic development, but instruments made that way are no less personal. Even in modern Western societies, JH gatherings promote private musicking—mostly solo playing, without electronic amplification, and submerging into a meditative state (Morgan, 2017). For Siberian vocal and instrumental traditions such musicking constitutes the norm (Zabolotskaya, 2009).

Despite the JH's tuning to a certain fundamental frequency (**FF**), indigenous techniques obstruct standard Western notation (frequency-based), since the production of discrete pitches constitutes only a fraction of the possibilities (Alekseyev, 1991b).

**Example-4**

"Mary had a little lamb," produced on khomus by Erkin Alekseyev. This music demonstrates inauthentic, Western-style treatment of JH as a "frequency-based" musical instrument whose purpose is to accurately reproduce the pitches of "tunes" (http://chirb.it/HNcBsq).

The bulk of the JH's autochthonous repertory consists of speech-like articulations and special sound effects (Alekseev, 1988). Common among players, the verbal characteristics of JHs' sounds are all timbre-oriented (Zagretdinov, 1997). Each timbre-distinguished device usually carries a particular semantic value (Shishigin, 1995)[15].

The JH *tutti* would mess up the clarity of articulation, preventing the parsing of meaningful musical elements such as *tabygyr* (staccato) (Alekseyev, 1986). Therefore, autochthonous styles are strictly soliloquy-based. A good example is the Yakut *syyia tardy*[16] ("moderate playing"), nicknamed "talking *khomus*"[17] (Grigoryan, 1957), which is closely related to the vocal *toyuk* (Shishigin, 1995). *Toyuk* (song) is the most ancient form of *dieretii yrya*—a smooth-flowing, drawn out improvisational singing style, based on a limited set of melodic intonations, and timbrally individualized (Alekseyev, 2016). *Syyia*'s closeness to *toyuk* enables the simultaneous singing and playing of a JH.

[12]During the second half of the twentieth century, a number of blacksmith masters in the USSR, especially Yakutia, attempted to enhance the construction of JH by maximizing the instrument's dynamic capacities to make JH more suitable for concert stage performance (Dyakonova, 2011). Modernized instruments substantially differ from traditional ones—to such extent that some experts consider them "not simply an augmented khomus, but a new instrument in terms of acoustics and art" (Kunanbayeva, 1987).

[13]Personal communication with N. Mamcheva.

[14]The notable exception are those cases where making of an instrument has to pass through a specific magic ritual, such as divination of the tambourine for shamanic use, which might involve up to 9 specialists and take a few days.

[15]For instance, the choice of *ysyakh* (the summer solstice celebration) as a subject by a khomus-player usually involves performance of the *okhuokhai* rhythmo-formula (festive round dance), the *crescendo* effect (as if gradually approaching the festival), *kuoregei* (imitation of a larch-call, associated with sunrise), *kebe* (cuckoo call, announcing the onset of summer), *tabygyr* (associated with the sound of the horse trot and droplets of melting icicles), and *uyuu* (articulation of words representing shouting of excited people and singing of dance-leaders) (Shishigin, 1995; Alekseev, 1988).

[16]*Syyia* in Yakut literally means "slowly, staidly," and *tardy*—"to yank, pull," as to pluck the JH's lamella. *Syyia tardy* signifies a particular playing technique, where silent mouth articulations as though speak out syllables while the plucked lamella keeps vibrating. Its technique is also used in other styles that require increased clarity of articulation (Zhirkova, 1991). This underlines the priority of clear diction, precluding simultaneous engagement of many JHs.

[17]"Khomus" ("Khamys" in archaic spelling) is a Yakut Jaw Harp. The alternative meaning of this word is "Scirpus"—grass-like plant that grows next to rivers and lakes (Pekarsky, 1959, p. 3293). Most probably, the original instrument was manufactured of this grass—similar to the way modern Nivkhi and Ulchi make their grass JHs (Mamcheva, 2012, p. 50).

**Example-5**

"*Tuluktan doborum*" ("Friendly bullfinch")—spontaneously improvised simultaneous singing and playing on khomus by Agrafena Ptitsyna from Megino-Kangalasskii district of Yakutia. The performer starts in the style of degeren (rhythm-oriented, metrically regular), but after 30 s switches to another style, *dieretii* (smooth, metrically free style), and thereafter keeps alternating between the two. Instrumental and vocal parts interfere with each other: khomus as though obstructs attempts of the voice to carry out a coherent song (http://chirb.it/2D3pJh).

*Yakuts do not consider the khomus to be frequency-based*. They evaluate its sound not as "pitch," but as "coloration." Therefore, concurrent singing-playing constitutes "coloring" the same song. All Yakut music is predominantly solo-vocal, where "vocal" is understood as "mouth-driven"—incorporating both vocals and JHs (Alekseyev, 1991a).

The inclusion of the JH in bands is uncommon in indigenous traditions. It is bound to the South: i.e., India and Indonesia (Morgan, 2008). In northern Eurasia, JH playing is usually solitary, and an additional player can participate only if he/she is a close relative[18]. The JH "combo" is a recent phenomenon[19], inspired by Western chamber music (Alekseyev, 1988, p. 185–7)[20]. Still, Siberian ensembles resemble the jazz combo by limiting *tutti* to only the start/finish of the music. Superimposition of JHs is understood here as a timbral mixture rather than as "frequency intervals"—unlike the counterpoint-style mentality of Austrian and German performers who group up to five or six JHs to produce "chords."

## TRANSMISSION AND PARTICIPATION IN PERFORMANCE OF TIMBRE-BASED MUSIC

Frequency- and timbre-based musics fundamentally differ from one another with respect to transmission. It is much harder to reproduce someone's timbre than their pitch. Timbre is usually personalized. We recognize a person's voice and a musical instrument by its timbre. Consequently, timbre-based music is designed to reflect the state of a unique individual rather than that of a group. Timbre (Saitis and Weinzierl, 2019) and sonority (McAdams, 2019) have been inherently connected to the display of emotions—even within the Western classical music tradition (Maddox, 2009).

- **Frequency-centered music** *attunes the individual to the collective* for their mutual benefit, enmeshing and empowering each participant through integration[21];
- **Timbre-centered music** attunes the individual to the surrounding nature, supernatural forces, a soulmate, or "oneself-in-the-past"[22]—all of which also integrate but do not enmesh.

The inherent individualism of timbre-centered music originates from mother-infant vocal communication. Their vocalizations imitate each other in pitch-contour and timbre to establish a personal bond (Malloch, 2000). These imitation games are deeply private, defining conventions of a "**mini-culture**" for each caretaker-infant pair (Trevarthen, 2008). Such a mini-culture opposes "culture" by circumnavigating social conventions for the sake of securing effective communication with any child, no matter how anomalous he/she may be.

Participation in a group always imposes obligations, and group music makes no exception. However, timbral music performance is fundamentally "free." The performer plays to his/her own satisfaction rather than the satisfaction of others. This is also true with a mini-culture: the caretaker and infant lock in a sympathetic communication that preserves individual freedom.

For timbral music cultures, infant **mini-cultures** develop into an adolescent "**maxi-culture**"—a maximally inclusive culture that allows everyone to remain himself/herself, rewarding them with a positive experience.

Timbral music acts like self-directed speech in problem-solving situations: it encourages behavioral reorganization and compensates for negative experience. And here the JH exemplifies timbral music. Notably, in many indigenous Eurasian cultures the JH is renowned for being a *female and children's* instrument (Mazepus and Galitskaya, 1997). The JH creates a mini-culture for each player within the ethnic maxi-culture.

## PERSONAL SONG (PS) AS A FORM OF TIMBRE-BASED VOCAL MUSIC

A peculiar form of "timbre-based" music is "**personal song**"[23] (**PS**)—a traditional custom of assigning a specific "tune" to each individual that represents his/her identity and satisfies the required sonic uniqueness: i.e., it does not resemble

[18]JH duets occur either during dating or in joined entertainment between two female friends. Thus, in Bashkortostan, leisure sessions might engage up to 3 women playing together (Shchurov, 1995).

[19]One of the first accounts of such practice comes from the time of collectivization campaigns in Middle Asia after the establishment of the Communist rule. Vinogradov (1938, p. 88) describes the performance of a JH orchestra of Kyrghyz young women as a "haphazard juxtaposition" of multiple instruments "without any coordination," where each performer played according to their own whims.

[20]The brightest features that distinguish this new use of JH from the autochthonous traditions of the past are its presentation from concert stage for a large audience, use of microphones and continuous concurrent playing of more than two instruments. Noteworthy is the globalization tendency of this new format—it attracts artists from many diverse ethnical backgrounds, including those that have not retained their autochthonous JH traditions (e.g., Russians).

[21]The empowering effect of collective frequency-centered performance is most evident in such genres as the work-song, military music and the national anthem—all called forth to increase group cohesion and to strengthen each of the participants.

[22]The propensity for self-communication – uniting self-reflection while performing, to the memory of oneself perceiving the same lyrical expression at some earlier time(s), was identified by Yury Lotman in regard to poetry (Lotman, 2001, p. 25). Musical expression can also follow suit. In fact, this is exceedingly common in traditional cultures of almost all indigenous people of Siberia (Sheikin, 2002, p. 304).

[23]This term originates from writings by Merriam (1964, p. 83). In English language literature, the term has been used in reference to "private songs" of North American Indians, e.g., Sioux (Austin, 1930), who, along with Canadian indigenous population, share the institute of PS with Siberian ethnoses (Ojamaa, 2002). In Russian research literature the most common equivalent term is "*lichnaya pesnia*."

the timbre and melody of another person's singing (Novik, 1999). A "personal tune" incorporates a particular melodic contour and timbre, and is used to coin multiple variations that the singer spontaneously forges while singing/humming during everyday activities (Novik, 2003). Most of one's day is accompanied by such semi-conscious singing. A comparison of different versions of one's PS recorded on different occasions shows great variability in text and emotional states along with stability in the melodic structure (Ojamaa and Ross, 2004), suggesting a link between "self" and melody. Furthermore, when a person is absent and missed by relatives, they will sing that person's PS as a substitute for his/her presence (Dobzhanskaya, 2017). The unauthorized performance of a PS by strangers is prohibited, and is punishable by the offended party (Grachiova, 1983, p. 56)[24].

Parents create PSs for their children as though "sound-painting," employing acoustic parameters to express the visual-motoric traits of the child (hyperactive/calm, headstrong/social, big/small).

**Examples-6/7**

Ex.6. Dinimiaku's children personal song, created and directed to Dinimiaku by her father, Tubiaku Kosterkin, from the settlement Ust'-Avam in Taimyr, in Nganasan. The song uses the timbral markers: contrasting "clean" high and "dirty" low registers connected together by sliding, and embellished with the "tremolo" effect. According to the performer, the song expresses tenderness and playful teasing. The lyrics question if the girl is upset at the old grouchy woman (Dobzhanskaya, 2014, p. 159) (http://chirb.it/7htOzK).

Ex.7. Derkuptie's children personal song, performed as if "coming from him" to the house guests, by his mother, Valentina Kosterkina, in Nganasan. The song is "infantilized": raised in pitch and "whiny"—it emphasizes the descending sliding intonations, as though gently "complaining." The lyrics imply that the infant-boy is overwhelmed by the visitors' attention and wishes to be left alone (Dobzhanskaya, 2014, p. 156) (http://chirb.it/zsxAEm).

After reaching adulthood, one might create or accept as a gift a second PS. Such renewal addresses the discrepancies between "infantile" and grownup states. The new PS often emulates the PS of a relative who is considered a role model[25].

**Example-8**

Adult personal song of Ver'a Nenyang, performed by his daughter, Liubov' Nenyang. Nenets masculine personal songs are characterized by praising one's own strength, luck, and smartness (Nenyang, 2006). Nenets adult personal songs usually oppose two contrasting vocal registers, joined by strong portamento, use of few "degrees" (3 in this example), and exaggerated intonations - the melodic intervals between the degrees are stretchable within the range of about 700 cents (Dobzhanskaya, 2017) (http://chirb.it/dgenGN).

The third[26] PS is permissible at old age to reflect age-induced personality changes (Sheikin, 1996, p. 67).

**Example-9**

*Murun Yrya* ("nasal" song). Old age personal song of Maria Sleptsova-Kustui Maaia, of Yakut-Evenk ancestry, reproduced by her granddaughter, Marina Vasilyeva. Her grandmother used to sing this particular arrangement during knitting. The lyrics list many different garments made by Maria in the past and explain what they are good for (Dyakonova, 2014). The song is characterized by nasalization and *kylysakh*, applied to the simple formula of 3 pitches ("degrees") (http://chirb.it/4Fc24f).

A PS resembles a renewable passport photograph, identifying principal temperamental shifts throughout life[27]. In this way a PS maintains *personal integrity,* reminding the singer of an anchor-state to return to after an emotional perturbation. This is most obvious in lengthy "autobiographical" PSs, where singers list dramatic events of their life (Ojamaa, 2002). A PS is replaced if the return to the anchor-state somehow becomes impossible.

PS "anchoring" is invaluable in the solitary lifestyle of Northern nomads, who are infamous for psycho-pathological disorders known as "arctic hysteria" (AH) (Tseng, 2003). In the colonial past, this diagnosis reflected the Eurocentric evaluation of the indigenous population by Western physicians (Tseng, 2007), but throughout the twentieth century AH captured the attention of anthropologists (Foulks, 1985). Eventually, AH fell within the scope of cross-cultural ethno-psychiatry as a "cultural syndrome" (Kirmayer, 2018). Such qualification somewhat undermines the life-threatening power of AH (Czaplicka, 1914, p. 307–25), perhaps because the English literature on AH focuses on Inuits, whose symptomatology appears milder and rarer than in the Russian North. Mitskevich (1929, p. 21) reported severe disabilities (and even deaths) in 60% of AH incidences in Upper Kolyma. He considered this to be an underestimate, however, since many concealed AH (13), and locals regarded it a "bliss" not to be medically treated (16). According to a Mitskevich's confidant, 100% of women in his settlement suffered from AH (21). AH ran through many families (24) and sometimes broke into mass "epidemies," affecting up to a third of the

[24]The PS taboo can be illustrated by the following case: after a newly installed system in the settlement's club-house played back a PS recording of a recently deceased local old woman, the residents of Chukchi settlement in dismay demanded that its public broadcasting be stopped immediately (Novik, 1999).

[25]One often imitates a relative's PS when moving away from home, due to suffering from nostalgia. Reminiscing about a dear relative by singing a similar PS on a daily basis, then, serves as combating nostalgia.

[26]In the past, for many, if not most, Siberian ethnoses, multiple personal songs were common. However, during the communist rule the institute of PS was severely shaken by campaigns directed against "shamanism," replacement of traditional lifestyles with modern ones, and the growing influence of "metropolitan" Russian culture, empowered through the education system. Thus, traditional taboo on non-personal use of PS was undermined by public broadcasting of those popular songs that are based on "personal motifs" of their creators, such as those by the distinguished Chukchi musician, Gennadii Pananto (Vensten-Tagrina, 2008).

[27]Russian ethnomusicologists often nickname PS a "musical passport." Passport in USSR differed from Western passports in that it indicated one's permanent residential address (*propiska*)—in addition to identifying a person's face, name, and family status—very much what the PS usually indicates.

population (25–29)[28]. Many feared contracting AH from others. Even in modern Russia, ethnographers are afraid to research AH-related matters[29].

Earlier researchers explained AH by the negative influence of extreme cold and Polar night (Novakovsky, 1924). Later research has identified such contributing physiological factors as calcium and vitamin D deficiencies (Wallace and Ackerman, 1960). Similar symptoms were observed among Southern neighbors (Mongols), along with phobias and "copycat" syndrome—attributed to a sufferer's excessive submissiveness (Aberle, 1952). Alcoholism was also considered a contributing factor (Foulks, 1985), although alcohol was exceedingly expensive and thus unaffordable to the indigenous Siberian population before 1917 (Mitskevich, 1929, p. 16). More convincing is Mitskevich's connection of AH with chronic starvation, constant stress, excessive pre-pubertal sexual activities, and sensory deprivation, which affect the Northern population much more than the Southern (45) and are environmental rather than cultural—they are as common among local Russians as among Yakuts (9). Such etiology and epidemiology are confirmed by numerous authors (Tokarsky, 1893; Gamov, 1894; Sieroszewski, 1896; Sakaki, 1903; Bogoras, 1910; Vitashevsky, 1911; Anuchin, 1926; Shreibler, 1927; Shirokogoroff, 1929; Shternberg, 1936; Petrov, 1960; Grigoryeva, 1996).

The reduction in AH must be attributed to the considerable improvement in living conditions throughout the twentieth century[30]. Prehistoric life in a tundra-like climate, then, must have been compromised by even more severe AH.

What remains unknown outside the Russian literature is the connection between AH and music. AH sufferers sing, first quietly, then more excitedly, swinging hands and shivering (Jochelson, 1910, p. 31). Their songs comprise a special genre—*menerik yryata* (crazy songs)[31]—characterized by a confusion of identity and consciousness: the singer haphazardly switches between multiple identities in dialogic singing, with incoherent words and intense timbral modulations (tremolo, rasping, falsetto), raving-like, occasionally shrieking, moaning, and clapping in metric disarray (Alekseyev, 2008). Sufferers report hearing songs of evil spirits, and can ameliorate their suffering by singing the evil spirits' songs, to vent the spirits out (Vitashevsky, 1911, p. 188).

**Examples-10/11**

Ex.10. Reproduction of *menerik yrya*, usually performed during the attacks of *meneriyi* by the anonymous old Yakut from Tattinskii ulus. He described his singing later as a reaction to seeing an evil spirit abaasy in the corner of his yurt and, causing him shriek in deep fear, trying to scare the spirit away. His yelling is interspersed with singing out the lines, supposedly pronounced by *abaasy*. Characteristic is the "dialogic" representation of at least two characters (http://chirb.it/tIandL).

Ex.11. Another *menerik yrya*, imitated by Vissarion Gavrilyev from the Maar settlement in Niurbinskii ulus - according to his experience of frequently witnessing *meneriya* of his neighbor, an old woman. The lyrics are more comprehensible than in the example above and present an argument between the patient and the spirit *ichchi*. The song is characterized by the alternations of a recitative-like excited singing/talking and brief tremolo motifs in a free metric setting, interrupted (rather than accompanied) by spontaneous clapping (http://chirb.it/mr28fk).

AH singing drastically contrasts with the stability of PS and its regular, endless repetitions of the same formula in a characteristic "personal" timbre. This opposition reveals the orderly power of PS and how crucial it is for survival in Arctic conditions. Continuous singing during long, solitary travel in the tundra prevents the rider from falling asleep and losing track of direction, or helps in surviving snowstorms (Krushanov, 1987, p. 234). PS keeps one's mind present under critical pressures.

In contrast, *menerik* loss of personal identity corresponds to suicidal behavior—e.g., running away and freezing to death (Nissen and Haggag, 1988)—supporting the indigenous beliefs that *meneriya* occurs when evil spirits invade one's soul. According to anamnesis, singing starts precisely during the attacks of such "possession" (Mitskevich, 1929). Although pathological singing clearly *results* from AH, there might be some *feedback* between the disruption of "normal" PS use and AH incidents. The Nanai, another PS ethnos, regard any mental disease as a "personality disorder" caused by the invasion of *multiple* spirits (Shimkevich, 1896)[32]. Culture is known to construct methods for handling common mental illnesses, which often include music (Robertson-DeCarbo, 1974). If music can heal, it should also have the power to aggravate—PS disfunction could contribute to AH.

[28]The documented evidence often indeed details the description of cases where witnesses who had observed the AH behavior shortly developed the same or very similar symptoms themselves. Sometimes these symptoms did not disappear by time, but gradually exacerbated into regular fits. This was especially common amongst females.

[29]Thus, Yevdokiya Sergina, a professor of the Arctic State Institute of Art and Culture in Yakutsk, in a personal communication, admitted to have deliberately avoided contact with people that suffered from AH disorders in fear of their detrimental influence. To illustrate the potential dangers involved, she reported the suicide of a young researcher who specialized in shamanic mysteries, after having attempted to investigate such disorders.

[30]The modern reduction in severity and incidence of AH could also be explained by the growing acceptance of AH by the local population, who tend to avoid medical assistance—in combination with the lack of qualified scholars capable of establishing rapport with local population to obtain confidential information. Most likely, the exuberance of the anamnesis information in nineteenth to twentieth century Russia became possible because of the custom of condemning political prisoners to settle in Siberia. Most physicians and ethnographers who left accounts of AH were convicted to exile at first by Imperial and later by Bolshevik courts—a long-lasting tradition that was instituted from the times of Russian conquest of Siberia (Hartley, 2014). The length of their sentences secured their bond with the indigenous people. Additionally, political convicts of the nineteenth century (as well as Bolshevik officials appointed to investigate local living conditions) used to be atheistic—unsusceptible to fear of evil spirits.

[31]*Menerik yryata* singing that includes moaning, shouting and occasional clapping with lyrics as though coming in alternation from presumable "evil spirits" and the singer himself/herself, is still encountered not only amongst Yakuts but also their neighbors, who have their own names for such singing: *cármoriel* amongst Yukagirs, *haujan* amongst Evenki (Jochelson, 1910, p. 31).

[32]Shimkevich describes a traditional Nanai wooden idol (*seon*) and an iconographic image drawn on a piece of cloth, paper or wood (*girki*) of a deity (*burkhan*) that has multiple heads, representing the disarrayed and conflicting mentality (p. 49). Nanai call such multi-headed *burkhans* "*segemi*." A similar tradition was described in neighboring Nivkh culture (Shrenk, 1903, p. 118).

Another musical-psychological anomaly, common in Siberia even now, is *tyyl yryata*[33]—sleep-singing, caused by extreme exhaustion or distress, especially when these are chronic. It resembles sleep-talking and can last for hours until the sleep-singer awakes (Jochelson, 1910, p. 13:37). Locals do not consider this a disease (Mitskevich, 1929, p. 10).

**Examples-12/13**

Ex.12. Genuine *tyyl yrya*, captured by Eduard Alekseyev from an overnight recording of Prokopii Sleptsov from the settlement Druzhina, Abyiskii ulus. Upon listening to this recording, Sleptsov remembered his dream of hunting a moose, but could not recognize the language of his singing (presumably, Yukaghir, a native tongue of his mother, that he later forgot). Singing is based on a brief descending gliding motif consisting of 2 pitches (degrees), possibly with a third complementary degree (http://chirb.it/086zkG).

Ex.13. Reproduction of *tyyl yrya* of an old woman, a relative of the famous Yakut singer, Luka Turnin, who had overheard her singing on numerous occasions. The song is based on the repetitions of a brief formula of 6 tones, engaging 3 pitches (degrees) - most likely a personal motif of the old woman. The lyrics complain about disappointing the barn-master spirit, and promise to please the spirit with a gift (http://chirb.it/cg0cvL).

The formulaic structure of *tyyl yrya* suggests that it constitutes a deeply ingrained PS, "automatically" reproduced by the sleeper. Perhaps *tyyl yrya* is a byproduct of the hyper-activation of the self-identifying circuit that attempts to cope with a stressful exertion, whereas *menerik yrya* reflects the failure of such coping. This would confirm Jochelson's observation that sleep-singing relates to *meneriya*.

In indigenous Yakut culture, there are numerous vocal genres dedicated to meditative psychological self-regulation in stressful situations: *kögus yryata* (pain-reducing songs), *enelgen yrya* (complaining songs), and *sulanyy yryata* (dying songs), which are collectively described as *uiulga yrya* (singing for mental endurance) (Dyakonova and Grigoryeva, 2017).

In Arctic conditions, the significance of PS is truly "Cartesian": "I *sing*; therefore I *am*" (Sheikin, 2002, p. 330). Loss of PS amounts to "disintegration."

Musically, a PS reflects one's affiliation with a kin and territory (Sheikin, 1996, p. 12). Specific intonations, rhythms, and timbres characterize specific geographic locations and kin groupings (Novik, 2004, p. 80). Comparative musicological analysis of PSs allows for identification of one's genealogical tree in Samoyed cultures (Niemi et al., 2004). The need for geographic tagging originates in the custom of taking a wife from neighboring ethnoses (Goltsova et al., 2005); territorial PS markers prevent incestuous marriage (Aizenshtadt, 1982). Patriarchic lineage determines the wife's social identification through her symbolic "rebirth," in which she loses membership in her birth kin upon adoption by her husband's kin (Sagalayev and Oktiabr'skaya, 1990, p. 18). Kin membership legitimizes a human being (Tyukhteneva, 2015). In such a system, the absence of PS amounts to "excommunication" from traditional society and the patronage of supernatural forces.

## THE CONCEPT OF MUSICALITY IN TIMBRE-BASED MUSIC

The existential cardinality of PS increases the value of musicality. In Western societies, a lack of musical abilities bears no biological cost; amusiacs do not suffer damage to their social life (Patel, 2010, p. 371–379). In PS cultures, however, amusia is costly.

Most traditional indigenous societies engage all members in active music production, since musical deficiencies pose serious obstacles to "normal" social interaction (van der Schyff, 2013). Like speech, music-making fuels **autopoiesis**—ongoing biological self-organization that optimizes the antithesis "I"-"non-I" to support one's autonomy (Maturana and Varela, 1980). Essentially, this is no different from the formative role of musical communication in the cognitive development of an infant (Cross, 1999). The absence of such communication endangers the psycho-emotional well-being of the child (Malloch and Trevarthen, 2009).

Inability to sing one's PS is perceived in indigenous societies as a handicap that requires family assistance. That is why, in Chukotka, those incapable of singing have their relatives sing their PS for them (Dyakonova, 2015). According to Siberian mythology, aphonia distinguishes dead souls from the living, and losing one's voice to a spirit equals death (L'vova et al., 1989, p. 90). Through its association with breathing, voice is considered an immanently live object capable of "invading" a person and "subduing" his/her mind (Pashina et al., 2005, p. 50). This is of special concern for shamans, who frequently imitate the voices and personal songs of dead people. Hence, using a "correct" voice is crucial for personal and societal well-being, and in some communities traditions call for the very old and sick to stay silent during public rituals (Dorokhova, 1995).

Unsurprisingly, in PS cultures the standard of vocal musical proficiency is decidedly low. Sieroszewski expressed discontent at "voiceless" Yakuts, who were always singing their "endlessly repeated monotonous tunes," "pleasant only to [the] performer," and "annoying like [a] mosquito buzz" (Sieroszewski, 1896, p. 569). His annoyance is evidence that contemporaneous Polish and Russian folksong standards, influenced by the classical *bel canto*, exceeded the Yakuts' in aesthetic demands. Frequency orientation and the absence of PS in Russian folksong enabled the *bel canto* influence, propping up its underlying idea of the necessity for "musical gift" and "proper" education. Such an attitude pushes toward *exclusiveness* of musicking and its professionalization. PS, on the contrary, promotes *inclusiveness*[34].

[33] *Tyyn yryata* ("night songs," in Yakut), or *tyyl yryata* ("sleeping songs") is unconscious singing in sleep, often sung in foreign tongues that the singer does not speak when in an alerted state. This is still relatively common amongst Northern Siberian ethnoses, especially men. Evenks call this singing *náyani*, and Yukaghirs—*yendójenut yáxtei* (Jochelson, 1910, p. 13:37).

[34] To illustrate this point, in 1957, one of the authors of this paper, Ivan Alekseev, irretrievably damaged his vocal cords by the excessive singing of his PS in nostalgia, after having moved away from his homeland. Nevertheless, his "bad" voice in no

Timbral music does not observe "wrong notes": informants are puzzled by questions about musical mistakes, as they believe that any expression is "right" (Ojamaa, 2003). In addition, intervals in timbral-based cultures are not fixed, as they are in frequency-based cultures, but are stretchable (*ekmelic*) depending on the singer's emotional state (Nikolsky, 2015). A performer, when repeating a song, sets the same lyrics to varying pitches, unaware of pitch discrepancies despite a fine musical ear that is evident whenever he/she sings other styles of music. Indefinite pitch certainly reduces to a minimum the demand on musical skills necessary for a "decent" performance.

A PS can be exceedingly simple, performable by virtually anyone.

**Example-14**

> Personal song of a Nenets old woman, Utchi, from the Kazym river region, covered with taiga (courtesy of Triinu Ojamaa). The melodic contour of this song's formula is exceedingly simple—engaging only 3 degrees within the range of only 274 cents: the lowest (228–231 Hz), the middle (249–252 Hz), and the highest (c. 254-267 Hz) degrees (http://chirb.it/0wM28B).

Structurally, a PS consists of "leitmotif" and "leittimbre"[35]—a reused motif and timbral quality. One family's PS can be distinguished from another's by its timbral style (Ojamaa, 2002). Musical elements that stay intact in all uses of a PS represent the PS's owner (musical "foreground"). Elements that change represent the circumstances of his activities ("background"): e.g., riding engages a different rhythm than fishing in an otherwise identical PS. Together, *foreground and background define the position of an individual in space and time*—indispensable in the solitary conditions and open spaces of Siberia.

Discrimination between foreground and background features requires the perception of ***thematicity***—musical material that constitutes a ***theme*** (Mazel, 1979, p. 150–5)[36]. Although thematic analysis evolved within Western classical music (Réti, 1951)[37], it should nevertheless be extended to all types of music (Val'kova, 1992)—like the Greek *thema* (proposition), which emerged in rhetoric theory and therefore suits any application related to semiotic representation where the perceiver needs to remember a particular expression essential for a musical work (Drabkin, 2001). Matters are complicated by the theme's salience. It might fall anywhere between two poles: "concentrated" (e.g., symphony) or "dispersed" (prelude) (Val'kova, 1992, p. 33)—as of the case in centonization in plainchant (Treitler, 2007). Timbral music also utilizes themes of various salience; JH players "compose" by developing selected thematic material[38].

PS culture relies on ***monothematicism***: "personal leitmotif-leittimbre" represents the same person. PS users identify themes intuitively, without conscious differentiation between melody and lyrics, but nevertheless *associate musical sameness with personality, and verbal diversity with environmental circumstances* (Ojamaa and Ross, 2004). Worth noting is that altering a PS's melodic contour constitutes a change of ownership between family members, but altering a PS's words does not (Ojamaa, 2002). A person sings one's PS without words when no contextual reference is needed, only a personal reference, such as missing a dear relative. However, changing words *per se* is considered changing the music. Typically, asking one to sing "another song" results in retexting the same **monotheme.** Only upon request to "sing like someone else" would a singer switch from that monotheme to the PS of the person mentioned (Alekseyev, 1988, p. 163–65).

"Timbral musicians" parse songs primarily by lyrics—not surprisingly, since phonemic oppositions constitute the timbral domain (Ojamaa and Ross, 2011)—rather than by pitch, like "frequency musicians" and non-musicians (Bonnel et al., 2001)[39]. Lyric-based parsing is common for Western infants (Lebedeva and Kuhl, 2010). Centering on lyrics generally characterizes the initial stage of acquisition of vocal musical skills (Welch, 1994). The earliest children's non-imitative singing surprisingly resembles an *ekmelic* PS with its stretchable intervals, formulaic structure with retexted lyrics, and private musicking to accompany various activities (Moog, 1976; Dowling, 1984; Bjørkvold, 1992; Campbell, 1998; Barrett, 2011; Koops, 2012). By 7 months, infants can differentiate the timbre of complex tones (Trehub et al., 1990) and remember timbre-specific information (Trainor et al., 2004).

If PS processing requires skills available to normal 1 to 2-year-old children, then amusia in timbral music societies

way affected the public judgment of his PS singing. This is in stark contrast to "bad" singing of Western classical compositions, which requires sustaining the pitch without fluctuations, retaining pleasant timbre, etc. (Sundberg, 1987).

[35] Although these terms originate from Western musicology, they are applicable to non-Western timbre-based music too. PS users reserve certain attributes of the melody and the timbre to make PS recognizable across all modifications. This is quite similar to what classical composers do in stage works (operas, ballets) and program music (tone poems and overtures). In both cases, the retained "leit-nucleus" marks the features of the "protagonist," whereas the changeable elements represent the "background" features, characterizing the context of action/happening.

[36] In Russian systematic musicology, the classic definition of "musical theme" is: "Musical idea that is fixed in musical structures, characterized by considerable completeness, salience, and originality—allowing one to recognize these structures upon repeated hearing—and that are often used continuously or reused within the same musical work" (Mazel, 1979, p. 151). Structurally, "thematicism" is the "totality of all themes employed in a work or in a group of works, which characterizes a particular music style (153). "Thematic material" is the complex of musical structures employed toward making a given musical work – disregarding their completeness and salience (152).

[37] The notion of "musical theme" is often described as peculiar to the Western classical music. This is not true. For instance, in South Indian Carnatic tradition, a musical composition is often segmented in 3 sections (*Pallavi*, *Anupallavi,* and *Caranam*) based on the treatment of the musical theme (Agrawal, 2018). The gamelan tradition not only utilizes the variation form, but also employs various forms of accompaniment to the principal theme (Miller and Williams, 2008).

[38] JH players choose a specific articulation (vowel or syllable), melodic intonation (a succession of pitches), imitation of a natural sound (birdcall) or a special technical device (staccato) for their specific expression. They use it either for repetition, variation, or contrast to some other thematic element – generating a sort of expressive "scenario" for a musical piece.

[39] Although Western listeners do integrate melodic and verbal information in lyrics prior to parsing song's lyrics, they process verbal and musical structures *separately* while listening to *familiar* songs (Sammler et al., 2010). In contrast, for indigenous Siberian listeners, autonomous processing of pitch and text does not seem to occur under *any* circumstances. Of course, final conclusion should be drawn only after experimental research of song's perception by representatives of well-preserved "timbral cultures," such as the nomadic Nganasan or Evenk.

must be virtually non-existent. Indeed, native PS-processing skills are widespread and effective. There are accounts of Siberian non-musicians retaining memory of a once- heard PS for decades (Novik, 2003). Timbral music culture is definitely designed to support this personal thematicism. Niemi et al. (2004) considers the absence of collective singing and instrumental pitch-based accompaniment among Northern autochthons a consequence of the identification functionality of PS[40].

## JAW HARP (JH) AS A PRINCIPAL MUSICAL INSTRUMENT OF EURASIAN TIMBRE-BASED MUSIC

Ps provides a model for the "musicalization" of other sounds. Animal calls, wind, rain, etc., all are regarded as the "personal voices" of natural objects, and are characterized by specific timbres and pitch contours (Novik, 1999). Like human PSs, their monotheme could take different shapes to provide information about their "whereabouts": e.g., Evens distinguish between the sad, angry, and happy "talk" of the fire in a fireplace. Moreover, "fire-talking" is believed to interact with human speech (i.e., respond to human conversation). Every environmental/household object could interact with humans. And this subjects them to the same "personalization" as humans.

Most Siberian **phono-instruments** (Sheikin, 2002, p. 46–67) possess their own "PS": flask, spoon, or cane, each is easily recognizable by its sound. Environmental sounds too represent "personal owner-spirits."

**Examples-15/16**

Ex.15. *Vyvko*—the Nenets buzzer used to imitate wind. In the past this had to do with the rituals of calling on rain, but now it is primarily a children toy, promptly made from a thread and a button (http://chirb.it/mcGarg).
Ex.16. *Symysky*—the Khakass male maral call, made from a piece of birch bark (http://chirb.it/8zt1tw).

To a human, such instrumental "PSs" form an "objective" sound-scene collectively representing the surroundings.

And, here, the JH emerges as the simplest archaic instrument capable of producing *multiple* "PSs" of natural objects—it is universally renowned for its onomatopoeia capacity (Maslov, 1911; Beliayev, 1933; Le Roux, 1950, p. 2:507–8; Picken, 1957, p. 186; Koizumi et al., 1977; Alekseyenko, 1988; Zagretdinov, 1997; Bulgakova, 2001; Sermier, 2002, p. 103; Alexeyev and Shishigin, 2004; Mamcheva, 2005; Canave-Dioquino et al., 2008; Yesipova et al., 2008; Suzukei, 2010; Kartomi, 2012, p. 159).

**Example-17**

*Temir-khomus* (Altaic JH) is imitating something that human voice cannot—the sound of the water stream, Tuva (http://chirb.it/aa35p2).

Onomatopoeic versatility explains JH's popularity amongst children, which reflects their desire (typical in traditional societies) to resemble adults[41].

Onomatopoeic vocalizations play an important role in early verbal acquisition (Menn and Vihman, 2011). Onomatopoeic words are easier to spot, remember, and comprehend (Laing, 2017). Similarly, onomatopoeic sounds facilitate comprehension of JH music, help in mastering the JH, and model new expressive means. The sounds one makes in learning the JH are no different than the babbling that comes from an infant learning to talk; both are playful explorations of the phonemic variants of a selected articulation (Davis and MacNeilage, 1995). The frame/content theory explains not only the acquisition of speech (Davis and Zajdo, 2008) but also of JH playing. If the vocal apparatus constitutes the infant's first "sound-toy" (Papoušek and Papoušek, 1995), for many timbral-based cultures the JH is the second. The "frame" of each JH syllable is filled with selective spectral content according to the same principles of phonetic symbolism that govern verbal acquisition (Shinohara and Kawahara, 2010).

The universal semantic values of ideophones and phonaesthemes (Svantesson, 2017), especially in vowels (Fischer-Jørgensen, 1978), apply to JH articulations. Thus, [i] implies nearness, narrowness, forwardness; [a] broadness, openness; front vowels tenderness, clarity, luminosity; back vowels firmness and concreteness[42]. Many associate sounds with colors/tints.

**Example-18**

The comparative demonstration of the principal khomus articulations: (a) front vs. (b) back, (c) high vs. (d) low vowels—performed by Ivan Alekseyev. Each of these 4 "poles" of JH articulations is characterized by salience of a particular register in generating a respective "JH formant": 2.4–3.1 kHz for "front," 0.3–1.6 kHz for "back," 1.2–2.5 kHz for "'high," and 0.6–0.9 kHz for "low" vowels. The most similar are the "back" and "low" vowels, distinguished by greater intensity of the lowest 1st, 2nd, 3rd, and 4th harmonics of the "back" vowel versus the "low" vowel which has much narrower bandwidth of its lowest formant (http://chirb.it/Az51G7).

"Front" syllables are described as tense and unpleasantly "tart."

**Example-19**

[40]In Siberia and Russian Far East, the exceptions from strict monophony are scarce: only the practice of collective repetition of the leader's verse by the chorus of dancers in Yakut *osuokhai* and similar dances of the neighboring peoples (Alekseyev, 1967), and, perhaps, the singing of the shaman's sidekicks during his *kamlaniye*—a ritual of reaching spirits necessary to fulfill the client's need (Dobzhanskaya, 2008).

[41]Thus, Khalkha Mongol women teach their children to play JH using imitations of horse and camel gaits. This is closely related to the herding family business in which children become involved at an early age (Pegg, 2001).

[42]Ivan Alekseev emphasizes that this generalization relies on the same synesthetic association of "smallness" with front vowels and "largeness" with back vowels that underlies phonetic symbolism of speech. JH articulations require constant control of all the constituents of the vocal apparatus, and reconfigurations of larynx and pharynx are usually done automatically—based on the reflexes established by speech.

"Front" syllables alone. They require tension in throat, face, lips, jaw, cheeks and tongue—isolating the mouth chamber. Erkin Alekseyev, a distinguished khomus player and the director of the research department at the International Museum of Jaw Harp at Yakutsk, describes his sensations while playing or hearing this articulation as though tasting an extremely sour apple (http://chirb.it/HNwHpq).

"Back" syllables feel pleasantly relaxed.

**Example-20**

"Back" syllables. They require relaxation in vocal apparatus. Erkin Alekseyev describes his sensations as "comfortable" to the extent of feeling "lazy" (http://chirb.it/MCzDkN).

"High" syllables resemble "smiling," from joyful to sarcastic.

**Example-21**

"High" syllables. They resemble "front" articulation, except that facial muscles remain relaxed. Erkin Alekseyev experiences this configuration as though "smiling" to oneself as in a situation when finding something funny. Emotionally, this state can be charged with joyfulness/playfulness (positive) or sarcasm (negative) – depending on how strained the larynx is (http://chirb.it/n7z749).

"Low" syllables are experienced as "sublime," yet constrained.

**Example-22**

"Low" syllables. They strongly activate the soft palate, configuring it into a "cupola." Majority of khomus performers associate it with the sound of a big church-bell—and imagine something sublime and lofty while playing or listening to it. Like "high syllables," sensing "low syllables" can appear "positive" or "negative"—depending on whether it is accompanied by the exertion of larynx and the discomfort resulting from this (http://chirb.it/6Iy1z8).

JH musicking explores a subject by chaining such elementary "meanings," interspersed with the referential meanings of onomatopoeic devices. In *syyia tardy*, expressing oneself via the JH is preferable to speaking/singing whenever one feels unclear about something and tries to figure out his/her attitude toward it. JH proficiency enables one to pick any thematic idea and track it, exploring one's emotional state—akin to self-directed speech, yet addressing subjects that are too intangible for verbal expression.

**Example-23**

Demonstration of a typical session of "playing for oneself" on Yakut khomus by Ivan Alekseyev (http://chirb.it/bg32m9).

Numerous North Asiatic ethnicities did not develop frequency-based musical instruments, probably because the JH satisfies the need for an objective reference frame and the PS for a subjective frame, thus leaving no void to fill.

## COMPLEMENTARITY OF PS AND JH PLAYING IN TIMBRE-BASED MUSIC CULTURES

The JH presents a counterpart to the PS. Onomatopoeia, citations of popular melodies, the use of conventional genre and stylistic features, and a "storytelling" compositional approach, common among JH players, all elaborate on environmental *objects*. Moreover, the JH is "anti-personal": it replaces the player's natural voice.

The JH's lamella makes women, men, the old, and the young sound the same despite all their physical differences. Its capacity to conceal the most telling source of personal information, the human voice, while supporting diverse mimicking, reflects the JH's power to "objectivize" sound. The camouflaging capacity of the JH goes a long way—it even allows aphoniacs to emit sounds (Shimomura, 2016).

Despite its "chameleonic" nature and different constructions, the JH remains easily recognizable (Ledang, 1972). Its "proprietary" timbre is evident in the practice of vocally imitating the JH with help from the fingers, a popular activity among Altai children (Fomin, 2018). Paradoxically, the "proprietary" timbre does not impede the imitation of other timbres; the JH makes an effective **vocoder** (Leipp, 1963). The lamella-generated tone provides a "carrier signal," and the vocal apparatus, a "modulator filter." What is unusual is the *hybridization* of different domains: the JH's modulator is human, while the carrier is instrumental.

- The JH is not an ordinary musical instrument that produces a desired tone, it is a musical "centaur": an organically indivisible human-instrument (Alekseyev, 1991b).

All JH constructions are "centauric": their instrumental component determines the frequency, while the vocal component—its amplification/attenuation (Dournon-Taurelle and Wright, 1978, p. 21). Since personalization comes largely from the vocal component, the JH becomes "impersonal." Totally unlike the PS, the JH hides the player's identity, whether in romantic serenading, shamanic rites, animistic hunting rituals, or children's pretend-games.

Vocoding generally relates to hiding. Electric vocoders were invented for ciphering military telecommunications to conceal the speaker's identity (Tompkins, 2010). In popular music, vocoders surfaced in urban genres that replaced a "humanistic" expression with a "robotic" one; they also penetrated other genres to conceal a singer's personal traits (gender, ethnos, class, age) that were perceived as potential vulnerabilities (Dickinson, 2001). Likewise, in animistic societies the JH reduces the risk of the player being identified and hurt while coming in touch with "spirits" (Popov, 1949, p. 265). In modern societies, the JH protects against eavesdropping (Morgan, 2008).

- The JH remedies the undesired ramifications of PS usage.

However, the JH offers a workaround for its camouflaging: it supports personalization through *pitch contours* rather than

*timbres*. In the Russian Far East, a player engages individualized patterns, each of which exposes the JH "voice" in its own way—often enabling the recognition of a particular player (Mamcheva, 2012, p. 223). This applies to other Siberian instruments as well. The Udege perceive typical instrumental patterns as the "personal songs" of a specific instrument (Sheikin, 1982). JH patterns constitute "motifs" of "timbral music," equivalent to PS motifs; the equivalence becomes apparent whenever the JH borrows thematic material from a song (Sheikin, 2002, p. 7).

**Examples-24/25**

Ex.24. *"Hyttya-hyttya, syrdyk kγmmγt"* (Summer is coming), Yakut folk song, sung by Fedora Gogoleva, describes a bright sun looking out from the skies and sending its warmth to mark the beginning of summer. The song is based on a simple 3-degree formula with regular dance-like rhythm in the *degeren* style. However, timbrally, music stays in the *tangalai yryata* style — "palatal singing," where tongue is abruptly pushed to the soft palate while taking a loud inspiration (http://chirb.it/vpnhv1).

Ex.25. *"Hyttya-hyttya, syrdyk kγmmγt"* improvisation on khomus, Fedora Gogoleva. The same performer takes the melodic formula of the song above (Ex.24) and elaborates its motifs/articulations (*naigryshi*), arranged to form a proprietary JH composition that is intended to express joy and happiness. The JH version loses the traits of "palatal singing," but keeps the *degeren* style rhythm (http://chirb.it/dp391O).

JH personalization must have emerged *after* its vocoding capacities became established. As archaic cultures lose their animistic ideology, musical instruments turn into producers of a specific musical arrangement, purposed for a specific application, thereby obtaining distinct semantic value and becoming "informative" about their creator (Novik, 1998). Thus, Mansi use instrumental patterns along with PS to display the birthplace/kin of Bear Festival participants (Tchernetsov, 1971, p. 110). Mansi and Khants conduct annual riverboat races where musician crew-members perform patterns representing the locations through which the boat passes (Sheikin, 1990, p. 8–9).

Each musical instrument embodies a particular gesture abstracted from the performing motion, plus the timbre abstracted from that instrument's sound—to form a new modus of musical thinking (Zemtsovsky, 1987). Like any instrument, a musical instrument is a technological tool for the (re)-production of a desired sound/texture[43]. To earn the reputation of a folk instrument for a certain ethnos, the instrument must execute a culturally important function (Stockmann, 1987).

Ainu, Ulch, Evenki, and Indonesian JHs are surprisingly similar in construction, sound, and technique (Duvan, 2003). So are Udmurt, Bashkir, Tatar, and Komi JHs of the Kama region (Aleksandrova, 2017). Nearly identical are the ancient Japanese and modern Nivkh and Karelian JHs (Wright, 2001). Intercultural contacts can hardly explain the commonalities among areas so remote from one another, as attempted by Fischer (1986, p. 156). A more likely explanation is the psycho-acoustic properties of the JH. There are four transcultural traits of JH music:

- Onomatopoeic imitations in "musicking for oneself," discussed above;
- Concealing the identity of a "serenader"[44] or his/her message from bystanders[45] (Hauser-Schäublin, 1995; McLean, 1999, p. 265; Hsu, 2001; Levin and Suzukei, 2006, p. 116–17; Canave-Dioquino et al., 2008; Kartomi et al., 2008; Matusky, 2008; Sam, 2008; Uchida and Catlin, 2008);
- "Transporting" shamans to the other world (Agapitov and Khangalov, 1885; Czaplicka, 1914; Vasilevich, 1969; Galaiskaya, 1973; Rouget, 1985, p. 126; Alekseyenko, 1988; Vainshtein, 1991; Suleimanova, 1994; Maskarinec, 1995, p. 192; Duvan, 2000; Solomonova, 2000; Pegg, 2001; Sarangerel, 2001; Yousof, 2004, p. 70–71; Canave-Dioquino et al., 2008; Honeychurch, 2015, p. 19; Chuluunbaatar, 2016; Tadagawa, 2017b);
- "Supernatural" protection (Popov, 1949, p. 265; Dyakonova, 1981; Dobzhanskaya, 1991; Pegg, 2001; Tadagawa, 2017b)[46].

Common to all four traits is the representation of *multiple* entities by a *single* JH "voice."

*The JH acts like a wild card, adopting different identities to benefit the player*. For hunters, the JH attracts prey by imitating their sound (Alekseyenko, 1988), or by pleasing the master-spirit of a particular place to get his favor (Galdanova, 1987). For sweethearts, the JH unites them by appropriating the other's articulations (Haid, 1999). For children, the JH presents unlimited make-believe impersonations (Yekeyeva, 2014). All these applications disregard the player's identity and focus on the impersonation

[43]Other than having a peculiar timbre, each instrument is characterized by a specific pattern of sounds convenient for playing on it. Thus, mandolin is characterized by plucking tremolo, guitar by strumming chords, and cimbalom by hammering passages—each of which can be successfully imitated by other instruments (as in orchestrations of Sicilian, Spanish, or Hungarian style music). In musicological literature such characteristic patterns are usually put under the umbrella of "musical texture" (Frayonov, 1981). Most popular musical instruments have developed "glossaries" of such patterns that allow to tell a musical composition that was conceived for one instrument from another (Kholopova, 1979).

[44]For example, maidens of the Bontoc tribe sing inside their houses in response to young men's JH playing outside, trying to guess the identity of the player (Canave-Dioquino et al., 2008, p. 437). Amongst Li people of Hainan, in contrary, girls use JH to camouflage their voices in their confession of love to young men (Hsu, 2001).

[45]The capacity of JH to faithfully reproduce vowels and some consonants of a language is often employed to deliver words and phrases at the closest distance, so that the more remote occasional witnesses would not be able to eavesdrop on an intimate conversation. Thus, in Hmong tradition JH not only imitates linguistic phonemes, but also engages tone intonations (Poss, 2012).

[46]The extension of traditional beliefs in JH talismanic protection can be seen in the modern-day use of JH music for therapeutic purposes by the local Siberian population. Such is the program of "Meditative-imagery Jaw Harp therapy" at the Goriachii Kliuch Sanatorium in Krasnodar administrative district, where about 60% of patients report improvement from such problems as panic attacks, posttraumatic stress, acute depression, sleep disorders and irritable bowel (Kalinichenko and Alekseyev, 2009).

act using the JH as a "fit-all" entity. The JH's vocoding nature empowers it.

In archaic cultures, the JH opposes the PS mainly through the implementation of privacy:

- PS aims at an **individual** by referencing his unique voice to him and his relatives/countrymen;
- The JH aims at the **environment** by reproducing and thereby appropriating its voices, attaining harmony through "centauric magic."

Self- and world-orientations complement each other. Not accidentally, the JH and PS share the same territory (**Figures 1**, **2**).

## ANIMISTIC ROOTS OF TIMBRE-BASED MUSIC

Initially, the JH worked like a talisman, connecting its owner with a supernatural force via "magic contract" (Novik, 1998). The first JHs were most likely made from grass, e.g., the Nivkh *konga-chnyr* (Mamcheva, 2005); tree twigs, such as the Altaian/Tuvan *taia/yyash-khomus* (Modorov and Dvornikov, 2015); and splinters, such as the Yakut *mas-khomus* (Sheikin, 2002, p. 119). It is exceedingly easy to make such instruments, even for a child; the materials are gathered rather than manufactured. The antiquity of JHs can be deduced from myths ascribing the invention of the JH to a bear—a totemic ancestor for numerous Siberian ethnoses (Startsev, 2017)[47].

The JH might have originated from the "aeolian harp"—its eerie sounds produced by wind hitting a splintered tree (Duvan, 2003). Siberians believe that lightning purifies trees that it strikes, and items made of their wood have protective power (Suzukei, 1991). Additional protection came from totemic phytomorphism: thus, Nivkhi drew their ancestry from larch, Ulchi from cedar, Oroks from birch, and Ainu from fir (Duvan, 2003). Plants were believed to breathe and, therefore, to possess a soul, possibly to even have PSs, just like animals (L'vova et al., 1989, p. 89). *Live creatures could have PS.*

**Example-26**

> "Personal song" of reindeer Urdy, performed "on behalf of him" by his owner, Valentina Kosterkina, in Nganasan. The lyrics describe reindeer's exhaustion from work, complaints on a dog that likes to bite his legs and anticipation of a good rest (Dobzhanskaya 2014, p. 124) (http://chirb.it/M46DrK).

Personalized songs and instrumental patterns are also assigned to spirits to attract their attention, and even musically reflect the mythological family relations between spirits (Gemuyev and Sagalayev, 1986, p. 68). "Spiritual" Ugric PSs were transcribed by Sheikin (1990). They were used to call a patronizing spirit for healing (Voldina, 2017).

Ancestor cults dominated most of Asia. Like mountains, rivers, and animals, plants too were believed superior to humans and could be totemic ancestors. Even today, Khakassian and Altaic kin groups consider larch and birch their ancestors (Sagalayev and Oktiabr'skaya, 1990, p. 50–59). Myths tell of trees nourishing or giving birth to human kin-founders (Tadysheva, 2018). A dying nearby tree is seen as prediction of death for someone from a corresponding kin: for Jyc, fir, Kuzen pine, Komdosh birch, Tubalar aspen, and Todosh honeysuckle (Kypchakova, 2006). Cutting ancestral trees is still taboo in Altai. Trees were actually used to trace genealogical "trees": thus, *Irkit* kin drew their origin from father-honeysuckle and mother-birch (Potanin, 1883, p. 7).

Zoo/phytomorphic ancestors are depicted on the Uralic metallic disks, 1st millennium BCE, from the Khanty-Mansi Museum (Gemuyev, 2000, p. 63–64); one of them depicts a "plant-woman," identified by ethnographers as mythological *Por*'s ancestress, birthed by a bear after eating *poryg* (Heraclium sibiricum) (Tchernetsov, 1971, p. 91–93). *Poryg* and similar herbs are commonly used for making musical instruments in Siberia (Sheikin, 2002, p. 4–7); these are sacred for *Por* people. Such instruments probably served as kin talismans before becoming regional musical instruments (Vasilyev, 2016). *Kongon* (Leymus), used for grass JH-making, was also mythologically anthropomorphized into a woman (Kreinovich, 1928, p. 192), and was possibly ancestral to some kin.

Trees serve as kin markers for South Siberian Turks (Sagalayev and Oktiabr'skaya, 1990, p. 43), who distinguish an individual by membership in *seok*[48] (kin) and believe that the bones of co-members are made of the same "wood" type. Following the tree/forest paradigm, each person corresponds to some tree, and his kin to the same tree-species—in the forest that represents the entire ethnos (53–54). The correspondence of people-grouping to tree-grouping is also observed among Ugric ethnicities (57). In the 2010 census, 82 *seok*s were registered in the Altai Republic (Tyukhteneva, 2015). Standard Altaic identification includes a totemic animal, tree, and mountain (Tadysheva, 2016).

Within this identification system, a plant-made JH would represent an individual, kin, or birthplace by the instrument's "voice." Initially purely timbral, such identification eventually became "melodic," engaging personal patterns. This development likely came about through vocoding. Since a wife is considered

[47] Thus, Yenisei ethnicities consider JH to be the instrument of Kaigus'—a bear-like deity, master of all animals, who uses JH made of birch splinter to imitate voices of all animals and thereby exercises power to dictate to them (Sheikin, 2002, p. 126). According to myths, in order to secure prey, Kaigus' taught human hunters to play JH prior to hunting.

[48] The word "*seok*" in languages of Altaians literally means "kin," "generation," "bone," and has been used in reference to a specific form of societal exogamic patrilineal organization (Verbitsky, 1865). *Seok*'s meaning of "kin" relates to its secondary meaning of "bone" through a more general meaning of "remains" (cemetery), where "bone" is understood as "the quintessence of live matter, capable of securing future births"—a trans-Siberian convention (Sagalayev and Oktiabr'skaya, 1990, p. 39). Hence, "bone" is something that is left from the deceased ancestor and that ties him/her with posterity: native Siberians believe that members of the same kin share the same bone ingredients (40).




**FIGURE 1 |** Russian Ural, Siberia, and Far East administrative regions where JH is the principal musical instrument. Such regions are colored beige. This figure displays the area of dominance of the indigenous Eurasian JH traditions according to the current ethnographic data. This area is confined to Russian territory. The names of those ethnicities that favor JH in the scarce musical instrumentarium are colored blue—altogether, 21 ethnicities. Each ethnicity name in the map reflects the areas of that ethnicity's greatest concentration. The only region where the importance of JH tradition remains on par with indigenous ethnicities of Russian Ural, Siberia and Far East (blue names) is Austronesia. Orange color marks those ethnicities that are either losing the JH tradition altogether (e.g., Chukchi) or retain it but of diminished overall significance due to great versatility of folk musical instruments (Tatars). The territories outside of Russia where JH music continues to occupy an important cultural role (approximating those of the orange-marked names of Russian ethnicities) include neighboring Kazakhstan, Uzbekistan, Kyrgyzstan, Turkmenistan, Mongolia, Tajikistan, Northern Afghanistan, Northwestern China and Hokkaido (not marked on this map). The names for JH in indigenous languages are listed in English by Bakx and Crane (2017) and in Russian, by Galaiskaya (1987), and in the encyclopedia of musical instruments (Yesipova et al., 2008).

estranged to her husband's kin even after marriage, and is tabooed to call her husband's totemic entities by name[49] (Tadysheva, 2018), articulating names on a JH or onomatopoeically hinting at them would circumnavigate the taboo. This could explain the female affiliation of the JH among many Ural-Altaic ethnicities. As a rule, men married foreign women (Pakendorf, 2007), who then became restrained by the taboo. Most archaeological finds of JHs in Ural vicinities come from personal belongings in female burials (Aleksandrova, 2017). Across Eurasia, the JH is one of few artifacts typical for burials, indicating the JH's connection to the supernatural and ancestral (Oleszczak et al., 2018). In the Far East, the JH is still used in lamentations for deceased ancestors (Sheikin, 2002, p. 364).

The ubiquity of the JH in Asia can be explained by its shamanic use (Emsheimer, 1986). *Shamanism constitutes a unified religious system in the entire Uralo-Siberian area* (Alekseyev, 1992). Samoyedic female shamans still use a JH instead of a tambourine (Alekseyenko, 1988), the primary shamanic instrument across North Asia. Around 2000 BCE, the entire area from Bactria to China was strongly influenced by shamanism (Francfort, 1994).

In Altai, novice shamans are restricted to playing the JH until they achieve spiritual maturity (for some never achievable) and receiving a tambourine—a pattern reflecting the antiquity of the JH in shamanic practice (Vainshtein, 1991, p. 273). In Mongolia, shamans use a JH to initiate shamanic rituals and engage the tambourine only upon reaching a trance-state (Chuluunbaatar, 2016). This suggests that the JH's onomatopoeic and vocoding capacities make it more socially inclusive than the shamanic tambourine, as they do not require exceptional "supernatural" powers from the JH's user (Sheikin, 2002, p. 69–134). Surviving beliefs in the personal protecting capacities of the JH reveal its origin as an egalitarian "magic" tool, preceding the professionalization of

[49] A stranger uttering names of a person, his relatives, sacred objects of his kin, including sacred trees, is believed to potentially harm that kin. This taboo has imposed heavy restrictions on wives, who despite bearing their husbands' children, remained estranged from their husbands' kin—thus, until the XX century, an Altaic widow did not receive inheritance after her husband's death (unlike her daughter) (Yenchinov, 2009).




**FIGURE 2 |** Russian Ural, Siberia and Far East administrative regions where PS is still practiced. Such regions are colored pink. The names of those ethnicities that have preserved PS well are colored blue. Ethnicities that are losing their PS tradition are marked orange. The tradition of PS retains its social significance (blue names) only on the territories of modern Russian Federation. Outside of Russia, this tradition survives amongst Amerindian tribes of the northeastern US and Canada (Ojamaa, 2002) (not marked on this map). This tradition existed amongst Ainu (Sheikin, 1996; Novik, 1999). The indigenous names for PS are: northern Mansi *eryg*, southern Mansi *erei*, Enets *barae*, Nganasan *baly*, Evenk *ikaee* (singer's PS) and *alma* (PS of close relatives or friends), Amur Nivkh *lu*, Sakhalin Nivkh *lund*, Ainu *yayan yuukara*, Chukchi *chinitkin grep*, Koryak *sinkin ulikul* (nominal song), and *ikoleyavan kuli* (kin-song) (Sheikin, 1996), Kerek *kuligul*, Even *tiinmei* (singer's PS), *alma* (someone else's PS) and *ikaan* ("cover" PS that has turned into a popular song), northern Khanty *ar*, central Khanty *arae*, eastern Khanty *lulpany*, Nenets *khybants* (Sheikin, 2002). Comparison of **Figure 1** and this figure demonstrates that PS and JH are shared by Mansi, Khanty, Selkups, Kets, Yughs, Kamasins, Evens, Dolgans, Evenks, Yakuts, Udege, Yukagirs, Chukchi, Kereks, Chuvans, Koryaks, Itelmens, Nivkhs, Ulchi, Nanai, Negidals and possibly Ainu. Of them all, Mansi, Khanty, Evens, Udege, Nivkhs, Itelmens and Koryaks have retained both traditions well. Area of their traditional habitat—the Urals, the Arctic Circle, Chukotka, Kamchatka, Okhotsk Coast, Sakhalin and Primorye—must delineate the stronghold of "timbre-oriented music." Prior to the Russian colonization and Chinese cultural influences, which have both been spreading out the "frequency-oriented music," this stronghold most probably included also Eastern, Western and Southern Siberia, where today both traditions, of JH and PS, are relatively weak.

shamanism and the establishment of the tambourine as one of its attributes.

JH musicking surpasses PS in its accessibility. Across northeastern Asia, toothless people use a hammer or ax to play JH (Tadagawa, 2017b).[50] In Yakutia, the ax and JH are even welded together[51]. The JH is used as an aid in *alalia* (Everstova et al., 2019). Complete aphoniacs play the JH (Shimomura, 2016)—and probably have been doing so for a long time in the Far East. The usage of JH-like devices (*koouchyntzyy*) to aid in aphonia was recorded during the Song dynasty (Picken, 1957). Such "prosthetics" highlights JH's great importance in native cultures.

Together, a JH and a PS provide "axes of coordinates" in the animistic worldview. The PS defines the "subjective" aspect as follows (Sheikin, 2002, p. 255):

1) "Thus I sing" (individual);
2) "Thus my countrymen sing" (territorial);
3) "Thus my family sings" (genealogical).

These hierarchic levels also reflect an evolutionary progression, inferred from cross-examination of all existing Siberian traditions. Sheikin adopts the Even culture for a model, because it has retained all three

[50] Video recording of M. Angin, the elder of the village Auri, Ulchi district, playing JH with the help of an axe can be seen in the documentary film "Ulchi," made in 1967 by the Moscow Studio of Documentary Films (Duvan, 2003). This technique requires superimposition of axe's blade on the ring of a metallic JH. The same performance technique was also utilized by Nivkhs in the beginning of the twentieth century (Mamcheva, 2012, p. 54).

[51] Yakut master Piotr Krivoshapkin from the settlement, Khomustaakh (Namskii ulus), used to manufacture the JH-axe hybrid—which was filmed by a group of Czech filmmakers in 1967.

levels (263–8)[52]. In the past, Chukchi observed a similar hierarchy[53].

An analogous three-stage model applies to the JH, defining its "objective" aspect:

1) "This JH protects me in my environmental encounters" (individual);
2) "These onomatopoeic JH imitations control their original owners from my environment" (territorial);
3) "This JH style optimizes my kin's relationship with the environment" (genealogical).

This model might apply to other Siberian mouth- and phono-instruments.

The intra-cultural interaction of the PS and the JH promotes the emergence of JH patterns. Thus, Nivkh players often engage the same pattern for different compositions (Mamcheva, 2012, p. 222).

**Examples-27/28**

Ex.27. Nivkh JH PN by Vera Khein, as presented in her original improvisation which she named "The flatfish dance." This piece, like many others that were recorded by Natalia Mamcheva from Vera Khein's, is based on her personal motif consisting of two descending intervals: of a 3rd and a 2nd (e.g., E-C-Bb), where the latter is often marked by the shorter rhythm of the middle tone (C) and its multiple repetitions, marking the lowest tone (Mamcheva, 2012, p. 297–300) (http://chirb.it/aIrCvG).

Ex.28. Nivkh JH PN by V.M. Persina, as presented in her improvisation (No. 130 from Appendix 1, Mamcheva, 2012, p. 301). It uses 2 motifs: of an ascending 2nd (passing C-D-E or auxiliary C-D-C) and of an ascending auxiliary 3rd (E-G-E), where C and E are marked as anchors, and the middle tones (D and E) are often given shorter rhythmic values (http://chirb.it/2L65H3).

Nivkhs also differentiate JH patterns by specific timbral words matched to a genre[54] (Sheikin, 2002, p. 132).

**Example-29**

The Nivkh rhythmic formula "Kan Vai," typical for various instruments and vocal music, in JH music is characterized by the onomatopoeic dog-like articulation of "khav-khav" (Mamcheva, 2012, p. 119). This formula distinguishes the genre of a dog-racing music, used in festive competitions that are held by Nivkhi, Ulchi, and Negidals. This genre is also used during the sacrifice of a dog in an annual bear festival (103) (http://chirb.it/q2xxbJ).

The amalgamation of non-onomatopoeic JH devices must have accompanied the historic development of PS: timbral themes prototyped **timbral words**. Many indigenous Siberian cultures employ specific vocables[55] as ethnic/territorial markers (Sheikin, 2002, p. 251). Their timbral and rhythmic makeup could direct JH improvisation toward elaborating standards. Thus, Yakut *khomusists* frequently engage the timbral vocable "*hyttya*." Timbral words comprise "timbral phrases," employed as in magic incantations (Tchernetsov, 1987, p. 35)[56]. Animistic hierarchy somehow transforms into syntactic hierarchy.

As the traditional lifestyle modernized, animism and totemism lost relevance, along with the need for anonymity. Subsequently, playing the JH acquired melodic/rhythmic/timbral patterns idiosyncratic for a player. They differ from a PS by engaging a *repertory of playing devices* rather than a single *monothematic* device.

- The JH was preferable to the PS whenever timbral music was "about" an **external** (to the performer) "object" rather than an **internalized** "subject[57].
- The JH's **plurality** matched an "objective" perspective, while the PS's **singularity** matched a "subjective" perspective with its single "self" (for a mentally healthy individual).

## JAW HARP ARTICULATIONS VS. SINGING AND SPEECH

Musical and verbal vocalizations adopt opposite parsing strategies (Alekseyev, 1993):

- Comprehensibility of **elements.** Speech prioritizes the clarity of contrasts to mark discrete elements in the speech-flow. **Division** prevails through timbral **contrasts** (oppositions).
- Comprehensibility of **gestalt-total** (melody). Singing prioritizes the integration of sounds into a melodic line to secure an empathic response to it. **Unification** prevails through timbral **similarities** (equalization).

[52]Structurally, Even classification is based on the relation between music and words. *Tiinmei* ("to speak to oneself") is an improvisatory song, created strictly for oneself, using timbral words mixed with spontaneous prosaic mundane texts. *Ikaan* ("song") can be performed for others. It uses cliché fragments of poetic speech which, however, do not comprise poetic verses. *Alma* ("recollection") is a "cover-song"—it reproduces someone else's words and/or musical themes and is retained as a ready-made "plot" in the memory of the following generations. *Alma* consists mostly of poetic speech.

[53]Chukchi classification is based on social grouping. *Chinitkin grep* conveys one's personal identity ("I, who sings my song"), *roiyr'yn grep* reflects one's membership in the family ("We, who share a home, sing as I do), *chygrymngyat grep* presents the entire clan of related families ("We, who share fire, sing songs to ancestors as I do) (Sheikin, 2018, p. 257, 41). The latter type has disappeared in the mid-twentieth century.

[54]JH personal instrumental pattern for dog-riding imitates barking with short tones articulated on syllables "*khai t'ok khai-nak-nak*" or "*khavr-khavr khaf-khaf-khaf*," while JH pattern for self-entertainment is based on sustained monotones on "*ko lyo-lyo-lyo*," with vibrato of the lower lip or tip of the tongue.

[55]For example, all melodies of Ude songs include words "*yava-yava*," whereas Nivkh songs—"*anga-anga*" (Sheikin, 2002, p. 250–1). Such words differ from "regular" meaningful words by being verbally meaningless and conveying semantic information through phonic symbolism alone. Often, they require a specific rhythm and melodic contour, thereby exercising a formative influence on the melody.

[56]Thus, in his diary from 1926 to 1927, Tchernetsov cites the meaningless sentence "*Gav-ri-ke, gav-ri-ke, vas pyrish, vas pyrish!*" that he heard continuously repeated on the JH (*tumran*) by a Northern Mansi girl in a remote Sos'va settlement.

[57]For example, a *syyia tardy* can be dedicated to such subject as representation of a young man and can present an "aesthetic" contemplation on such a "topic." JH player envisages a protagonist (here, a young man) and fantasizes a storyline, more or less cohesive and tangible, that is not always describable in words. Such a composition is "about someone else" ("objective")—as opposed to the typical PS which is "about the singer" himself/herself ("subjective").

The JH's vocalization introduces tonal homogeneity and monotony. Both are considered detrimental to speaking and singing, reducing the intelligibility of their prosody. JH prosody is unaffected by this. *It relies on contrasts only of the mid/high-frequency spectrum, confining monotony to FF.*

In frequency-oriented cultures, singing requires timbral uniformity while stressing the FF changes (Titze, 1988). Voicing every consecutive utterance by *definite pitch per se* homogenizes the spectrum of sung tones[58]. Like singing, verbal intonation engages FF changes even in non-tonal languages (Ladd, 1996), while stressing spectral contrasts, as JH does.

- The JH **combines singing and verbal approaches**. Its monotony conceals the player's identity yet preserves all articulations: "timbral words" become monotonously "flattened." Such "flattening" presents a form of timbral equalization similar to singing.

Both singing and speech vowels stress FF (**Figure 3**). But speech's clarity depends more on frequency **changes** of F1 and F2 (Zsiga, 2013) that are sufficient for distinguishing vowels (Carlson et al., 1970). Singing **freezes** F1 at about the same frequency level for all vowels (Sundberg, 1974).

- Singers intuitively match vowels. Speakers oppose them. So do JH players—but only in relation to the "**residual** tone," not the "**fundamental** tone."

The concept of "residual pitch" was introduced by Schouten (1940) and adopted by modern theories of harmonic perception (Goldstein, 1973; Terhardt et al., 1982; Moore et al., 1984, 1985, 1986; Houtsma and Smurzynski, 1990; Moore, 2012; Norman-Haignere et al., 2013). The sum of the unresolved partials that comprise "pitch residue" is perceived as a pitch identical to FF, but harsher in timbre (Schouten et al., 1962).

Unresolvability does not prevent pitch perception, because pitch can be analyzed by an alternative mechanism: through the autocorrelation of nervous impulses, activated by partials, in a phase (Licklider, 1951). Despite its complexity, such hearing is exceedingly common (Moore, 2012, p. 217). However, unlike FF recognition, it requires learning (Terhardt, 1974). Hence, periodicity recognition comprises a cultural phenomenon.

- JH prosody engages monotony to facilitate residual pitch analysis[59].
- A drone-like FF turns into "background," while changes in residue comprise "foreground" in an auditory scene analysis of JH music. Its users focus on the foreground and ignore the background.
- As in a PS, the unchangeable JH "background" represents the hidden "self." The changeable "foreground" represents the "world."

JH users distinguish one articulation from another most likely by learning and memorizing a number of "harmonic templates," pairing them with the perceived sounds in a search for the best match (Goldstein, 1973). Such templates are formed by a long-term accumulation of statistical data, and tend to rely on the harmonic series (Shamma and Klein, 2000).

Musical templates might be influenced by phonological templates that are generated by the pitch analysis of the tonal attributes of speech (Schwartz et al., 2003). JHs made of organic materials usually deviate from the harmonic series in the tuning of partials. A JH can also omit a harmonic (usually f2), or generate a non-serial partial or a subharmonic (see Appendix-1 in **Supplementary Material**).

- JH's "centauric" nature makes JH's prosody **hybrid** (Alekseyev, 1991b).

The **equalized** FF portion of the JH spectrum resembles the singing formant (Sundberg, 1974), while JH's "residue" portion, in contrast, marks every **opposition** of phonemes. Hence, the *JH can "talk" AND "sing"* **sequentially** as well as **simultaneously**. JH prosody engages two different domains, vocal and instrumental, where the vocal controls the dynamics and the instrumental the frequency of the harmonic components (Trias, 2010). Each domain introduces its own formant(s) (**Figure 3**):

1) F1 is **instrumental**; it homogenizes JH articulations and depends mostly on JH's construction.
2) F2 is **vocal**; it differentiates between JH articulations and depends mostly on the vocal tract configuration. Additional vocal formants are present at least in "high" vowels [i:].

JH vowels retain F1 at the same level (per instrument), but move vocal formants. Each vowel has its unique harmonic signature comprised of:

- the number of harmonics over the noise floor ("active" harmonics),
- their dynamic envelope within the spectrum,
- an odd/even harmonic ratio,
- the harmonic composition of formants.

The border between the "instrumental" and "vocal" ranges lies ≈f6–f8. Therefore, we must introduce a new concept of "**harmonic base**" to encompass f1–f4(f7) in contrast to "**harmonic residue**."

[58]We must underline that the homogeneous nature of singing manifests itself *only* in frequency-based musical traditions. Vocalizations of timbre-based cultures, including PS and vocal "games," such as Inuit *katajjaq* or Ainu *rekkukara* (Nattiez, 1983), abide by principles of timbral contrast, and are considerably richer than conventional frequency-based vocal music. Such timbrally-rich vocalization styles are used by Chukchi, Nganasan, Koryak, Kerek, Even, Enets, Yukagir, Nanai, Oroch, and Ude (Sheikin, 1996). As Sheikin demonstrates in his analysis of these styles, their TO is based on the binary opposition of two "timbral classes": inspiratory and expiratory rasping, each of which is varied by employing different syllables and intonations. There can be as many of 20 "timbral classes" (e.g., the *pilche'ingen* vocal system of Chukchi).

[59]That is why multiple attempts to modernize the traditional JH constructions by adding extra lamellas or joining a few JHs on a single "JH farm" did not really catch up with majority of JH players. Western European players of multi-pitched JHs subscribe to the model of Western tonality borrowed from classical music. Chinese players follow the model of traditional Chinese music system based on *yayue*. Like Western tonality, this constitutes a type of **tonality**—frequency-based music (as opposed to timbre-based indigenous JH traditions).

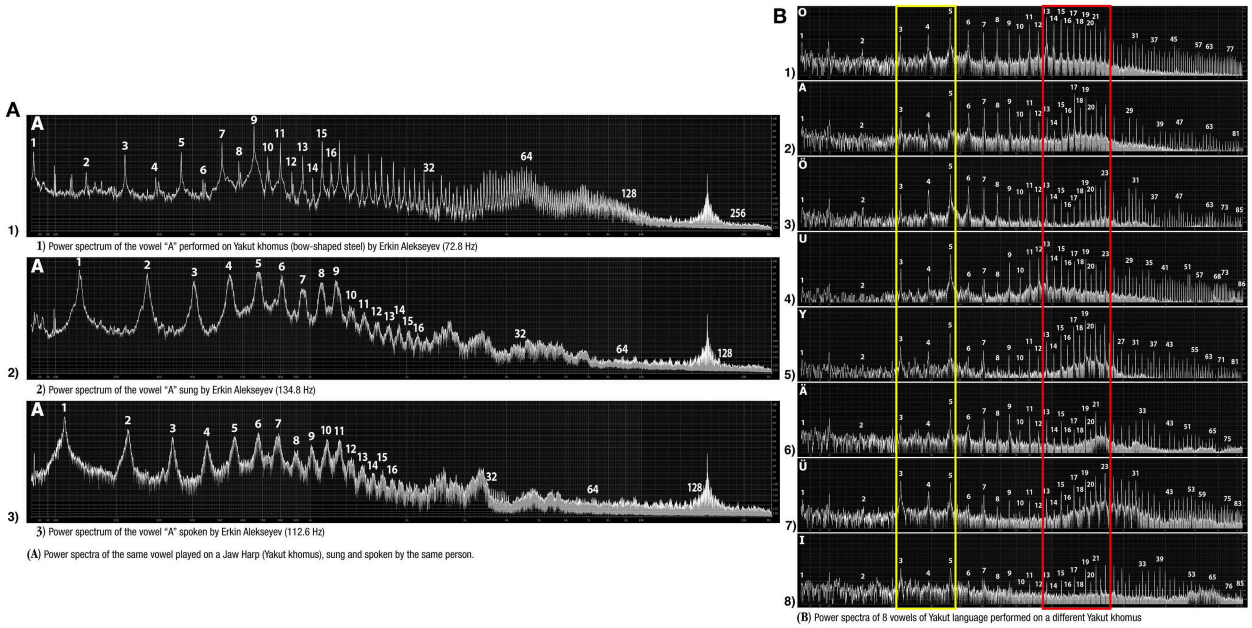



**FIGURE 3 |** Characteristic acoustic features of the JH vowel articulations. This figure demonstrates the importance of harmonic structure for the JH vocal system. The numbers on the curves mark the harmonics. Vertical axis indicates amplitude, and horizontal frequency. **(A)** Power spectra of the vowel [a] performed on khomus, sung and spoken. (1) Khomus. There are 145 harmonics discernible above the noise floor, peaking at f9. (2) Singing (bass). There are 23 active harmonics, peaking at f1. (3) Speaking. There are 19 active harmonics, peaking at f1. Weakness of formants seems to be part of the indigenous Yakut prosody—judging by the recordings of native Yakut speakers we had at our disposal. In their number of active harmonics, position of the loudest harmonic and distribution of peaks (2) and (3) are harmonically closer to each other than to (1). JH reserves higher intensity for odd harmonics, as opposed to speaking and singing. The bandwidth of harmonics is significantly narrower for khomus than for singing and speaking. If to measure the bandwidth at the harmonic's baseline level, the khomus f2 occupies a range between 141 and 151 Hz, making 117 cents (a semitone). The f2 of the singing "A" ranges between 243 and 297 Hz (349 cents—a neutral third). And the f2 of the speaking "A" ranges between 197 and 238 Hz (325 cents—a narrower third). The last notable difference is a much smaller noisiness of the khomus "A"—easily detectable in the absence of the "furry" looking jiggles in the contour line. These jiggles are very prominent in the singing of "A" and even more so in the speaking of "A," especially >f7. The recordings were made on the recorder Ritmix RR-989, with the sampling rate of 44.1 kHz and the bit depth of 16 bit, and analyzed by the software application RX Pro by iZotope with the following settings: 262144 samples FFT size, Hann window, 4x time overlap, average channel mode, 80 dB/s decay, extended log frequency scale. The peak at 15.6 kHz reflects a permanent background noise in the room. The analysis was the average of 4 versions of (1), (2) and (3). **(B)** Power spectra of "active harmonics" of all Yakut vowels articulated on Yakut khomus. The vowels are numbered according to their perceived "pitch values" in an ascending scale (Ogotoyev, 1988), according to the convention common amongst Yakut JH players—which generally coincides with the comparative vowel's height. JH's "instrumental formant," F1, marked by the yellow rectangle, falls on f5 in all vowels except "Ü" and "I," where F1 broadens to f3–f5. In contrast, "vocal formant," F2, marked by the red rectangle, falls on different harmonics in 6 vowels: f13 for "O," f17 for "A," f23 for "Ö," f19 for "Y," f21 for "Ä," and f21-f23 for "I." Only 2 vowels, "U" and "Ü," engage the same harmonics as, respectively, "O" and "Ö." However, "U" differs from "O" by a smoother shape and consistent dynamic prevalence of odd over even harmonics (>f13). "Ü" differs from "Ö" by an even greater contrast of odd/even harmonics and louder high harmonics >f47. Altogether, the harmonic composition of F2 varies by the number of active harmonics, the ratio of the intensity of odd to even harmonics and the magnitude of dynamic "dropouts" amongst the adjacent harmonics in the harmonic series. Therefore, each vowel possesses its unique harmonic configuration of F2. The recording was made as in **(A)**. The audio clip representing each vowel was selected out of a pool of 15 performances by 3 performers on 3 JHs based on averaging the dynamic values for the lowest 11 harmonics and finding the closest match to the average harmonic profile for each of the vowels.

JH formants differ from speech/singing formants for the same vowels (Leipp, 1967). A JH vowel engages six times more "active" (i.e., salient) harmonics than singing, and almost eight times more than speaking. F1 in the JH is much softer than in speaking and singing. The JH's F2 is much stronger. The JH's harmonics display a much greater odd/even contrast than singing/speaking vowels do. JH harmonics are much narrower in bandwidth and are compromised by less noise.

**Harmonicity** (Plack, 2010), which has no direct parallel to the JH for singing and is absent altogether from speech, is definitely crucial for JH prosody (Nikolsky et al., 2017). Very likely, harmonicity/inharmonicity forms the principal axis in the discrimination of JH articulations (Nikolsky, 2017). Indigenous users must have high sensitivity to the harmonic structure of sounds. Thus, Nanai believe that everything sung will certainly be heard by spirits who do not recognize human speech (Bulgakova, 2000). The harmonicity of singing could be responsible for this.

The hybrid nature of JH prosody poses the question: is it derivative of speech, like a grammelot (Jaffe-Berg, 2001); of singing, like a kazoo; or is it ancestral to both language and music?[60]

[60]This matter is thoroughly examined in the acoustic and ethnographic study that is scheduled for publication by the state People of the World Khomus Museum and Center in Yakutsk, Russia next year.

## ESTABLISHING THE TIME AND PLACE OF ORIGIN OF INDIGENOUS EURASIAN JH MUSIC TRADITIONS

Archaeological evidence points to Inner Mongolia as JH's homeland (**Figure 4**).

The oldest JH belongs to the Lower Xiajiadian culture, 2146–1029 BCE (Kolltveit, 2016). Its proximity to the Manchurian sites where the Neolithic "singing masks" were found (Okladnikov, 1971) indicates that JH articulations might have originated from timbral singing (**Figure 5**). "Singing masks" always stress the configuration of mouth, nose, and eyes in a single face, omitting the body; in contrast, the animalistic art of the neighboring Siberian cultures presents the entire animal, and usually in groups (Okladnikov and Mazin, 1979, p. 86). Each "mask" depicts a specific vocal articulation charged with an emotional expression that may represent an ancestral attribute (Sheikin, 2002, p. 259).

Modern Nanai tell myths about a Skull Horse-rider and paint burial idols' faces to house dead souls, closely resembling skull-like Sikachi-Alyan petroglyphs (Okladnikov, 1979). The Nanai habitat includes Manchuria, and the JH remains their principal instrument. "Singing masks" could capture the symbolic meaning of a particular articulation and assign it to an ancestral figure as its "call"—like an animal call ascribed to a totemic ancestor. The model for this could be military calls, *urany,* used by Turkic peoples to distinguish a kin (e.g., Yakut "*urui*!") up until the nineteenth century (Sagalayev and Oktiabr'skaya, 1990, p. 21). Such calls were believed to possess supernatural power, making them suitable for the vocoding "protection."

Vowel harmony characterizes most languages of the Altaic Sprachbund (Ko et al., 2014), especially of the Turkic language family (Gadzhiyeva, 1997). There, the Yakut vocal system (Antonov, 1997) is considered the most ancient, representative of proto-Turkic (Tenishev, 1984) and proto-Altaic languages (Tenishev, 1997). The acknowledged paleolinguistic studies date the break of proto-Altaic by 8000–10000 (Andreyev and Sunik, 1982), 7000 (Starostin, 1991), or 5000 BCE (Robbeets, 2015, p. 506)—earlier than the earliest of the Sikachi-Alyan images (Okladnikov, 1971, p. 83–89). "Singing masks," then, could capture the articulation representing a totemically important meaningful "harmonic" word, such as "*urui*"[61].

"Singing masks" spread over Amur/Ussuri, Baikal, Lena, and Aldan, dated 3000–100 BCE (Okladnikov and Zaporozhskaya, 1972; Okladnikov, 1974; Devlet, 1976, 1980; Leontyev, 1978; Arkhipov, 1989). This distribution mostly coincides with the oldest archaeological JH finds and modern areas of greatest JH popularity. The northern Far East became culturally isolated around the Bronze Age and retained Neolithic traits until the seventeenth century Russian colonization (Andreyeva, 1987), effectively conserving the archaic traditions.

"Singing masks" are also depicted on Malyshevo ceramics, 2000 BCE (Okladnikov, 1968, p. 50). Similar Neolithic stone and clay masks were excavated in Primorye (Brodianskii, 1978). In Yenisei burials, near Sayan, "real" gypsum masks were found, manufactured ≈2000 BP, probably following some earlier tradition of mask-making from easily decomposing materials (Devlet and Devlet, 2011, p. 310–12). Nenets, Khanty, Mansi, Evenk, Udege, Kumandin, Shor, and Buryat shamanic masks likely represent the same tradition.

Closely related are *tsam* masks (335–37). *Tsam* is a theatrical ritual dance, performed in Lamaistic monasteries across Buryatia, Tuva, Tibet, Nepal, and Bhutan. To distinguish the tsam protagonists from each other, their masks express different emotions, assisted by melodies/rhythms that are specific for each protagonist. Shamans too accompany mask-wearing with vocal and theatric impersonations of evil or good spirits (Zabolotskaya, 2011). Instead of masks, Tuvan shamans use maskoids—small, mask-like plates, mounted on a hat or knitted, called "human face" (Devlet and Devlet, 2011, p. 313). Maskoids usually reproduce conventional face-masks and represent the ancestral spirits guarding the maskoid's owner (Avdeyev, 1960).

The mask's protective function (Ivanov, 1975) resembles JH's "masking": both are employed by shamans to repel evil spirits.

"Singing masks" likely belong to the cultural Neolithic/Mesolithic "confederacy," stretching from the Volga-Ural region to Tibet and China (Leontyev et al., 2006). Its influence may have extended farther to Southeast Asia and Oceania (Okladnikov, 1968, p. 52–64). *In this entire area the bamboo/wooden JH remains the principal musical instrument.*

Archaeological data confirms the vegetative origin of the JH. All BCE finds, except for Perm's, occurred in Asia (**Table 1**). The JH crossed the Urals into the Volga planes ≈200 BCE (Aleksandrova, 2017) thereafter heading toward Hungary and Ukraine. Finno-Ugric peoples might have carried the JH to Northern Europe.

Archaeological finds form 6 clusters: Mongolia->China, Mongolia->Baikal, Baikal->Ural/Volga, Volga->Ukraine, Manchuria->Japan, and Western Europe. Bamboo/bone JHs comprise two of the oldest clusters. Ural/Volga and Manchuria/Japan combine both organic and metallic JHs. European JHs are exclusively metallic. This clustering is exceedingly important, because metallic and organic JHs significantly differ in **spectral texture**.

## SPECTRAL TEXTURE AND ITS DEPENDENCE ON TECHNOLOGICAL EVOLUTION

The concept of musical texture (Skrebkova-Filatova, 1985) has not been applied to JH music. However, JH players definitely arrange spectral components in a way similar to parts in polyphonic music. Polyphonic thinking is common in a great many indigenous music cultures (Jordania, 2006), timbral music included. Tuvan vocal "solo-polyphony"

[61] In Yakut language exclamation "*urui*," especially common for shamanic verses, is used in prayers for well-being, repelling evil, expressing a jolly feeling, calling for fighting, or announcing an end to something (Pekarsky, 1959, p. 3070).

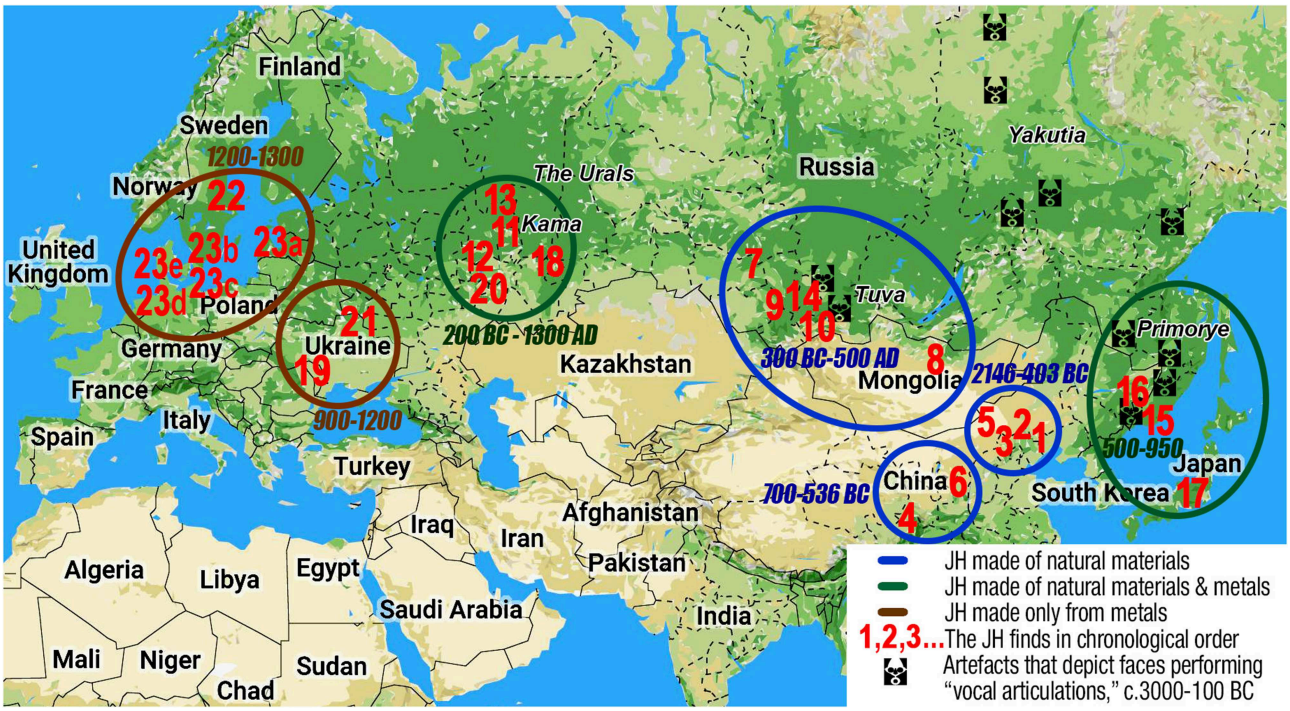



**FIGURE 4 |** Worldwide geographic distribution of archaeological finds of JH, dated prior to the fourteenth century, and of artifacts that depict faces performing "vocal articulations," dated c.3000–100 BCE. This map is designed to reflect JH's distribution pattern throughout Eurasia prior to the fourteenth century—the beginning of JH mass production, trade inclusion and, subsequently, export to European colonies. Numbers in red reflect the chronological order of JH finds. Their dating and publication sources are referenced in **Table 1**. Circles mark the geographic clustering of finds and are accompanied by their dating range. Blue color marks JHs made of natural organic materials (bone, bamboo, wood), brown—of metallic JHs, and green—of mixed (organic and metallic). The most ancient finds cluster within the Inner Mongolia and are all organic. Next in timeline are Southern China and Altai, both of which also house organic JHs. Still next is the cluster covering Ural Mountains and Volga planes. It includes mixed materials. Two subsequent areas to the West, Ukraine and Northern Europe, feature exclusively metallic instruments. Manchuria and Japan form another cluster, concurrent with the Ukrainian, and feature mixed materials. The black face icon marks the archaeological sites containing petroglyphs, sculptures, masks or ceramics that depict human faces with open mouth—"singing masks" (Okladnikov, 1968, 1971; Okladnikov and Zaporozhskaya, 1972; Devlet, 1976, 1980; Okladnikov and Mazin, 1976; Brodianskii, 1978; Leontyev, 1978; Arkhipov, 1989; Kochmar, 1994; Devlet and Devlet, 2011)—exemplified in **Figure 5**. Most singing masks petroglyphs precede JH finds and concur with the Shuiquan JH in Inner Mongolia. The oldest petroglyphs are located at Sikachi-Alian, Amur, 3rd millennium BCE (Okladnikov, 1971, p. 83–89), and the most recent—the utmost northern location in Yakutia, at Olenek River (Arkhipov, 1989; Kochmar, 1994). Geographically, the locations of "singing masks" coincide with the JH finds in two regions: Tuva and Primorye, but chronologically precede the JHs excavated in those locations, suggesting that JH articulations grew out of timbral singing, probably of PS. Absence of "vocal articulations" in Inner Mongolia could be explained by the lack of attention of Chinese archeologists, not distinguishing such pictures from other images [in the same vein, Chinese archaeologists had difficulties recognizing JHs in the unearthed materials of a number of sites (Shulga, 2015; Kolltveit, 2016)]. On the other hand, absence of JH finds in Central and Southern Yakutia and Kamchatka, where images of "vocal articulations" were found, might be due to the wet climate that makes preservation of organic materials unlikely.

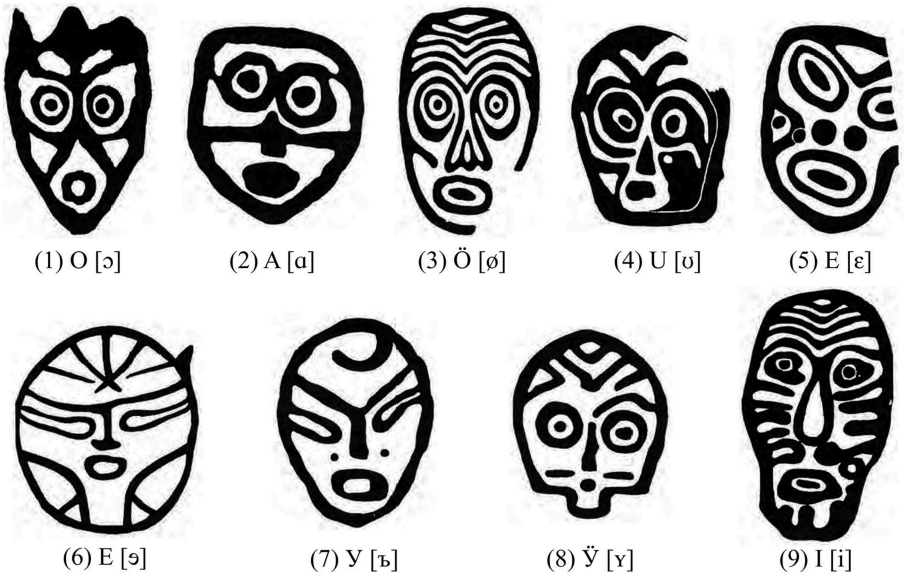
is well-known. JH traditions also engage 2- and even 3-part polyphony (Brodsky, 1972). Dobzhanskaya (1991) pioneered the polyphonic notation[62] of JH field recordings; an earlier attempt had been made by Pugh-Kitingan (1977) for a Huli JH excerpt. Polyphonic analysis must be used to investigate JH textural typology. It is hardly possible to comparatively study JH traditions without accounting for the distribution of thematic material within the spectral texture.

The JH's texture (i.e., the structure of the harmonic residue and harmonic base) depends on the instrument's material (**Figure 6**, see the archive "23 figures for Appendix," in **Supplementary Material**, for higher resolution images).

Grass and metallic bow-shaped JHs form two poles of textural arrangement. A grass JH generates **maximal polyphony**: five undifferentiated melodic parts moving simultaneously and only partially controlled[63]. A metallic JH generates **homophony**: a single melody with accompaniment. Bamboo/wood/bone

[62]She distinguished between 4 textural layers in the repertoire of Chuvan JH music for frame-shaped JHs: (1) the FF; (2) its subharmonic chest resonance an octave below; (3) the middle melodic layer above FF controlled by the player's breathing and tongue position, exhibiting "talking" qualities; and (4) a high melodic layer that changes the sustained phonemic formants which resembles singing.

[63]There is experimental evidence that non-musicians can simultaneously track no more than 3 concurrently running polyphonic parts, and trained musicians—up to 4 (Stoter et al., 2013). Although this was established in relation to frequency-based music, it can also apply to timbral-based music since most known traditions of vocal and instrumental timbral music too use no more than 3–4 parts. It is reasonable to expect the JH players to consistently and reliably control no more than 2–3 melodic parts. This assumption agrees with the anecdotal information from JH players.




**The vocal system reconstructed from these images:**

[ɑ], [ъ], [ɔ], [ʊ], [ɛ], [ɘ], [ɪ], [ø], [ʏ] – of which [ø] could be interpreted as [ɐ].

**Basic vowels from the phonetic alphabet after Nadeliayev:**

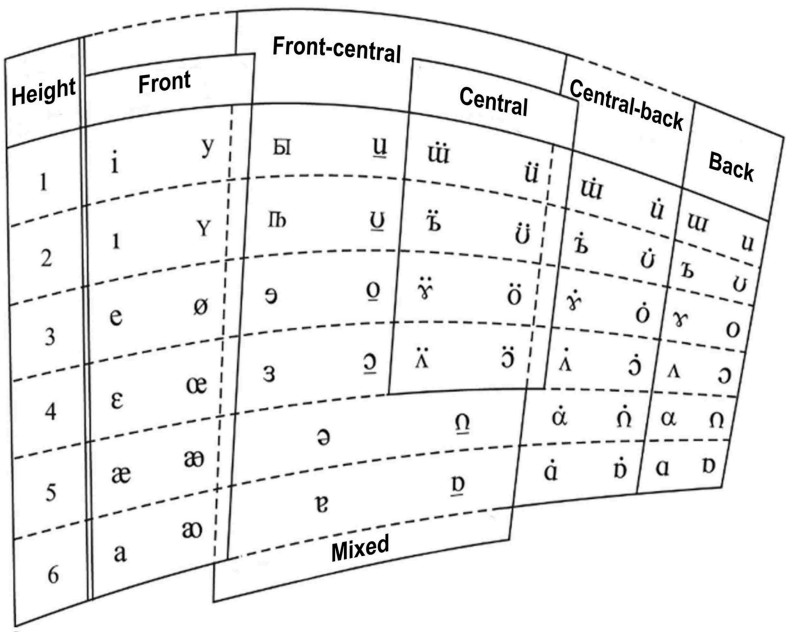



**FIGURE 5 |** "Singing masks" from Sikachi-Alyan, Amur river, dated c. 3000 BCE (Okladnikov, 1971, p. 83–89) and their phonetic interpretation. These 9 images were selected by Sheikin (2002, p. 259), who believed that they depicted pronunciation of vowels in order to perpetuate a model for articulating sacred names, important for the ancestor cult. Such images are found near Amur river along the border between modern Russia and China, and differ from most East-Northern Siberian petroglyphs by focusing on not animals, but humans. Specifically, these faces emphasize the configuration of open mouth and the emotional expression in an individualized manner where no one image reproduces any other image (Okladnikov, 1971, p. 85). Similar in style, "singing masks" are found in few isolated sites to the West of Primorye, in Tuva, and to its north, in Yakutia and Kamchatka, suggesting the enormous expansion of a Neolithic Far Eastern culture during the Bronze–Iron Ages (Andreyeva, 1987). Each mask could hint at a particular sacred word tabooed for public use—similar to the current practice in the Altai region. The vowel harmony, omnipresent in Uralic and Siberian languages (Kiyekbayev, 1972), especially strong in Turkic languages (Gadzhiyeva, 1997), in effect, encodes the configuration of the mouth for the entire word, since all its syllables are bound to reproduce the same vowel (e.g., Yakut miitary call "*urui*"). Sheikin's set of images is identified as pictorial representation of a vocal system of a hypothetical "Amuric" Paleo-Asiatic language by I. Seliutina, N. Urtegeshev, T. Ryzhykova and A. Dobrinina

*(Continued)*

**FIGURE 5 |** from the Laboratory of Experimental Phonetic Research at the Institute of Philology of the Siberian Department of the Russian Academy of Science. Based on the methodology of the founder of this institute, Nadeliayev (1960), this lab has collected an extensive database of articulations of basic phonemes of Siberian languages (Seliutina et al., 2012). Although the lack of uniformity in Neolithic images limits the researcher to estimate the gradations in labialization and only a gist of its palatability, however, Nadeliayev's classification of labialization (Nadeliayev, 1980) allows the phonologist to identify vowels based on the ratio of mouth width to height. The "Amuric" vocal system is presented in the phonetic alphabet developed after Nadeliayev by 9 of his pupils: I. Seliutina, A. Urtegeshev, A. Letiagin, A. Shevela, A. Dobrinina, G. Esenbayeva, A. Savelov, M. Rezakova, and Yu. Ganenko (Seliutina et al., 2012). The basic vowels from this alphabet are shown in a table at the bottom of the figure. The "singing masks" are numbered in the order of the ascending JH "articulatory scale" as formulated by Ogotoyev (1988), with the correction by Shishigin (2015) that places "o" rather than "a" at the bottom of the scale. The mask images (Sheikin, 2002) and the phonetic table (Seliutina et al., 2012) are reproduced by permission of the authors.

**TABLE 1 |** The earliest Jaw Harps found by archaeologists.

| # | JH type | Location | Country | Dating | References |
|---|---|---|---|---|---|
| 1 | 2 bone idioglot | Shuiquan, Jianping, Liaoning | China | 2146–1029 BCE | Kolltveit, 2016 |
| 2 | 1 bone idioglot | Xiajiadian, Chifeng | China | 1200–600 BCE | Honeychurch, 2015 |
| 3 | 4 bamboo idioglot | Yanqing, Beijing | China | 770–403 BCE | Tadagawa, 2016 |
| 4 | 5 bamboo idioglot | Yuhuangmiao, Sichuan | China | 700–600 BCE | Shulga, 2015 |
| 5 | 4 bone idioglot | Jundushan, Hebei | China | 700–500 BCE | Honeychurch, 2015 |
| 6 | 1 bamboo container | Baojiang, Shaanxi | China | 576–536 BCE | Tadagawa, 2017a |
| 7 | 1 bone idioglot | Dubrovinskii Borok, Kolyvanskii District, Ob' | Russia | 300–100 BCE | Borodovskii, 2007 |
| 8 | 1 bone idioglot | Morin Tolgoi, Altanbulag soum, Tuv aimag | Mongolia | 300 BCE-100 AD | Honeychurch, 2015 |
| 9 | 4 bone idioglot | Tcheremshanka and Tchultukov, Altai | Russia | 300 BCE-600 AD | Borodovskii, 2017 |
| 10 | 1 bone idioglot | Aimyrlyg XXXI, Tuva | Russia | 200 BCE | Tadagawa, 2016 |
| 11 | 1 bone idioglot | Makhoninskoye, Chastinskii district, Perm' | Russia | 200 BCE-300 AD | Aleksandrova, 2017 |
| 12 | 1 copper idioglot | Ust'-Brykinskii, Laishevsky district, Tatarstan | Russia | 300–500 | Aleksandrova, 2017 |
| 13 | 7 bone, 22 idioglot bronze | Prikamye burials and settlements, Udmurtia, Bashkortostan, Permskii, and Kirovskii regions | Russia | 300–1300 | Golubkova and Ivanov, 1997 |
| 14 | 2 bone idioglot | Sakhsar, Khakassia | Russia | 400–500 | Tadagawa, 2016 |
| 15 | 1 iron idioglot | Adrianov Kliuch, Partizanskii district, Primorye | Russia | 500–600 | Leshchenko and Prokopets, 2015 |
| 16 | 3 iron idioglot | Nikolayevskoye, Shaiginskoye and Smol'ninskoye, Mikhailovskii district, Primorye | Russia | 698–926 | Leshchenko and Prokopets, 2015 |
| 17 | 2 iron heteroglot | Hikawa shrine, Omiya, Saitama | Japan | 900–950 | Tadagawa, 2007 |
| 18 | 1 silver idioglot | Idelbayevckiye burial mounds, Salavatskii district, Bashkortostan | Russia | 900–1000 | Mazhitov, 1981 |
| 19 | 1 iron heteroglot | Echimauti, Rezina district | Moldova | 900–1000 | Feodorov, 1954 |
| 20 | 2 bone, 2 bronze idioglot | Tankeyevskii, Bulgur nekropol, Kuibyshevsky district, Tatarstan | Russia | 900–1100 | Aleksandrova, 2017 |
| 21 | 1 iron heteroglot | Hlukhiv, Sumi district | Ukraine | XIII century | Pashkovskii, 2012 |
| 22 | 1 iron heteroglot | Uppsala | Sweden | XIII century | Kolltveit, 2006 |
| 23 | 5 copper heteroglot | Riga, Skanor, Ribe, Greiswald, Hamburg | Latvia, Sweden, Germany | XIII–XIV centuries | Kolltveit, 2009 |

*The numbering reflects the chronological order from the oldest dates to the newest. "JH type" shows the number of JHs found in a given location, the material from which they were made and the type of construction they use. "Source" cites the publications that contain descriptions of each find.*

engage **well-differentiated polyphony**: bi-melodic for bamboo and tri-melodic for bone and wood (see Appendix in **Supplementary Material**).

The order of the increasing discretization of parts in these textures corresponds to the order of the growing complexity of manufacturing a JH from a given material (**Table 2**).

Materials fall into three classes according to texture, timbre, and playing technique (**Table 3**).

This order suggests the development toward simpler/clearer typologies where each textural type constitutes a stage that builds on a previous stage.

Obviously, the "metallic stage" comes after the "organic." Based on a comparative ethnomusicological perspective, Sheikin (2002, p. 132) offers further distinctions. He defines the "bamboo stage"[64] as representative of Oceanic cultures (exemplified in

[64]Sheikin speaks of the "bamboo period," but his choice of the term "period" is not the most accurate for designating the timeline of JH's evolution. The word "period" implies a "hard" end—one period replacing another. However, none of the JH materials have been really "discontinued" in the production of JH. Once mastered, a material usually remains in use. Therefore, the term "stage" seems more appropriate here, since developmental stages are cumulative in nature and built on top of one another.

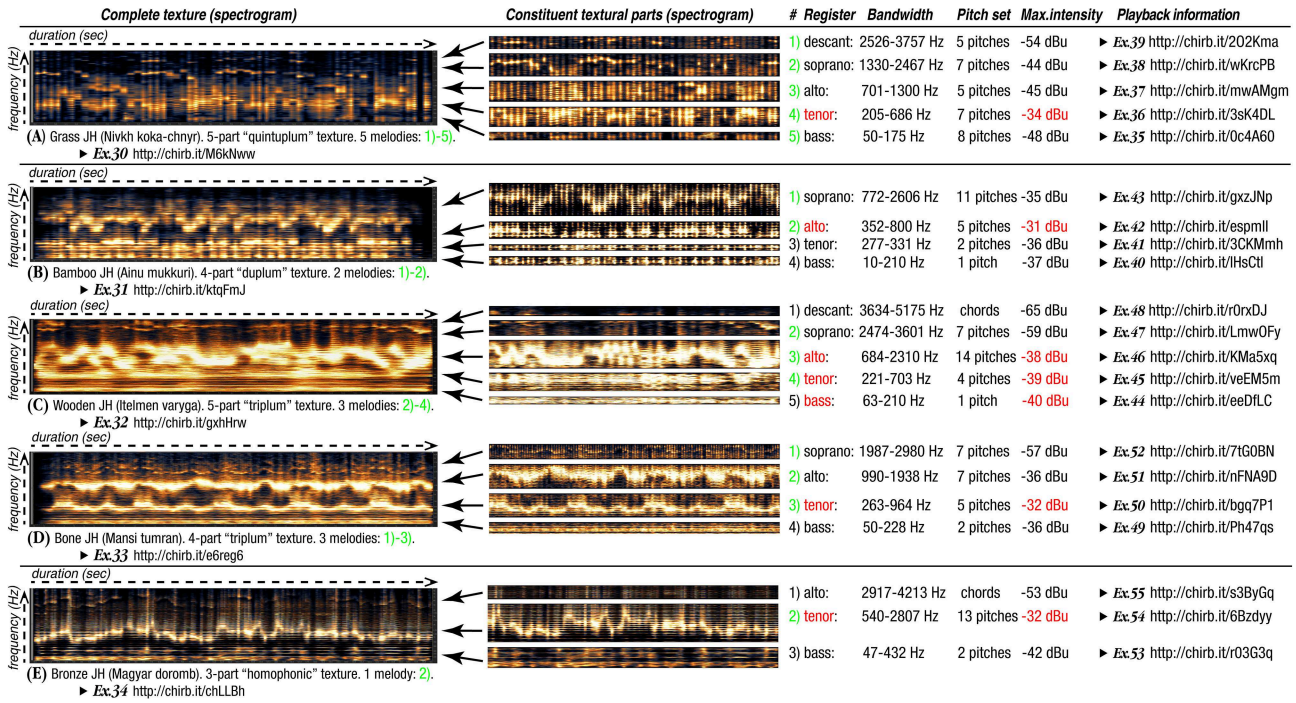



**FIGURE 6 |** Spectral textures of music performed on JHs made of 5 different materials. On the left, spectrograms show the spectra of JHs made of each of the 5 different materials, most commonly used in Northeastern Eurasia by indigenous players. The horizontal axis indicates time, the vertical axis frequency, and the color-coding amplitude (from black for silence to bright yellow for maximal intensity). In most cases, the texture spreads from about 70 Hz up to 3–5 kHz. On the right, the entire texture is broken in "parts" (bands), each capturing the frequency range within which the "active" harmonics form a continuous (as much as possible) melodic stream. The procedure for defining these parts is described in the Appendix in **Supplementary Material**. Arrows point to the original position of parts. To the right of the parts are their names, following the standard nomenclature of Western choral music. Each part's bandwidth, amplitude and number of pitch classes (i.e., pitch set) are listed on the right. A pitch class is defined by a dynamic spike repeatedly marking the same frequency level (plus/minus a quartertone) throughout the clip. The dBu value reflects the greatest amplitude for a given part of the texture within the entire clip. The red color marks the most intense part that presumably serves as the principal melodic line for the JH player. Whenever there are a few parts with many pitch classes that reach amplitudes that are close together, they are considered "polyphonic" melodies. The green color marks the "melodic" parts. Each of the complete textures and their constituent parts can be auditioned by downloading the corresponding audio file from the provided web link or looked up in the zip archive in the Supplement. For the description of Audio Examples Nos. 31-55, see the file "List of Audio Examples.docx" in the Supplements. **(A)** Grass JH (Nivkh). 5 Components, each melodic (including the bass). Each part uses similar number of pitches. The leading melody is in tenor. The other 4 melodies are of similar intensity, except the highest one. This texture can be called *scattered "quintuplum" polyphony*, since all of its parts are constantly moving in a similar way, making it hard to track their changes by ear. The only controlled part here is likely to be the tenor. **(B)** Bamboo JH (Ainu). 4 Components, of which the upper 2 are melodic. Bass sustains a pedal tone. Tenor is restricted to an ongoing figuration of 2 adjacent pitches ("principal" belonging to the harmonic series, and "auxiliary" that does not belong to the harmonic series but is a step away from the "principal" pitch). All parts are nearly equal dynamically. Soprano contains the most embellished hemitonic melody which uses twice more pitches than does the anhemitonic alto. This is a *well-differentiated "duplum" polyphony* of soprano and alto, with the drone accompaniment of tenor and bass. **(C)** Wooden JH (Itelmen). 5 Components, of which 3 middle ones are melodic. Bass sustains a pedal tone. Descant alternates between a few ostinato cluster-chords (one principal, others auxiliary). Three lowest parts are nearly equally intense. The polyphonic contrast is tonal (4 anhemitonic pitches for tenor vs. 14 chromatic pitches for alto). This is a *fairly differentiated "triplum" polyphony* of tenor, alto and soprano, accompanied by the pedal bass and pedal-like descant. **(D)** Bone JH (Mansi). 4 Components, of which the upper 3 are melodic. Dynamically, their relation resembles **(C)**, but the stronger tenor is likely to lead. However, tonally, there is little differentiation between 3 melodic parts. This is a poorly differentiated "triplum" polyphony, accompanied by the bass pedal. **(E)** Bronze JH (Magyar). 3 Components with a single melody in tenor. Tenor is the most dynamically and melodically salient part. Its melody is accompanied by the pedal bass and the ostinato "cluster-chords" in alto. This is a *well-differentiated homophony*. Black horizontal lines break 5 textures in 3 groups based on their structural similarity. **(A)** Stands out by its maximal polyphony and the absence of differentiation of parts. **(E)** Stands out by its absent polyphony and maximal differentiation of parts (pedal-melody-ostinato). **(B–D)** Have limited selective polyphony. Of the three, **(B)** features the strongest, whereas **(D)**—the weakest differentiation. **(C)** Features the greatest complexity. The complete spectrograms were calibrated to blur the display of the most intense parts of the spectrum in order to better indicate the melodic continuity across pauses, along the timeline. Blurring was achieved by maximizing the FFT size up to 65,536 samples, with the multiresolution STFT, no frequency overlap and time overlap 4x-16x, depending on the audio material. The window is Gauss 200 dB. Frequency is displayed in mels. The spectrogram of a complete texture zooms in from about 0–100 Hz to 4,000–9,000 Hz, depending on the audio material. Each clip was noise-reduced and dynamically normalized. The clips of textural "parts" have slightly increased frequency zoom and reduced blurriness (greater frequency and time overlap) to display finer detail. The audio of "parts," uploaded for audition, is normalized in order to make the parts comparable.

mukkuri) and preceding the "bone/wood stage" of Siberian peoples. Sheikin's periodization can be enhanced. Wooden JHs likely replaced bamboo as the JH spread north. Bone must have replaced wood in the further expansion to the tundra, where wood is rare. Hence the "bone stage" is likely to have succeeded the "wood stage."

Geometry/construction was another factor that determined the rise of homophony: only the bow-shaped metallic JHs are

homophonic. Nevertheless, both factors nearly always[65] coincide (Fox, 1994; Sheikin, 2002; Kolltveit, 2016). In contrast, frame-shaped idioglot JHs do not seem to support homophony in indigenous traditions. Therefore, *the adoption of bow-shaped metallic JHs ended up replacing JH polyphony with homophony.*

It is imperative to determine the timeline of the distribution of bow-shaped instruments (**Figure 7**).

Organic JHs emerged in Atlai/Baikal/Mongolia/Primorye during the pre-Metal Age and spread north, south, and west. Grass and chip probably succeeded twig, followed by chip-in-frame and frame-idioglottic instruments—corresponding to the advance from purely "sonoric" effects to the pitch-oriented manipulation of overtones (Sheikin, 2002, p. 124). An example of an intermediate "chip-in-frame" construction is the Tuvan *charty-komus* (Suzukei, 1989, p. 65–6).

In the Volga/Kama/Ural region the westward expansion of frame met a wave of eastward expansion of metallurgy to introduce metallic JHs (Aleksandrova, 2017). Initially, these seem to have imitated the popular organic constructions (Golubkova and Ivanov, 1997). By the Sarmatians' time, a single standard for frame-shaped instruments (10.3–12.5 × 1.4–1.7 × 0.1–0.2 cm) was established across Central Asia and southern Siberia (Borodovskii, 2017). *This must have corresponded to a single standard of polyphonic texture.*

Farther westward propagation of JHs involved only metallic instruments; wood/bone JHs were used in Europe exceedingly rarely and only before the heteroglot metallic constructions became established (Kolltveit, 2006, p. 83)[66]. The bow-shaped construction was invented somewhere in Ukraine/Balkans/Alps/Karelia, providing a new, simplified texture that co-existed with the older textural standard. However, the exclusive cultivation of bow-shaped JHs must have established a new textural standard that spread all over Europe and its overseas colonies[67]. On land, bow-shaped JHs spread eastward over the same territories where organic JHs had already settled, eventually reaching the Far East[68]. This brought about the co-existence of two autonomous traditions among Nivkhi, Ainu, Evenks, Evens, Kets, Yughs, Selkups, Chukchi, Itelmens, Koryaks, Kereks, Yukagirs, Khanty, Mansi, Tuvans, Yakuts, Dolgans (Sheikin, 2002, p. 125–6), Mongols (Pegg, 2001), and Chinese (Li, 1956). The two traditions differ in playing techniques, sound quality, and texture, to the extent of bearing different names among the same people (Li, 1956; Yakovlev, 2001; Sheikin, 2002; Mamcheva, 2012). The closer to the Far East, the greater the discrepancies. This geographic distinction is supported by the gender/age distinction. Frame-shaped JHs constitute the female/children's sphere of use—bow-shaped the male/adults' (Tadagawa, 2001; Sheikin, 2002; Dyakonova, 2017)[69].

Significantly, the traditions differ in their mythological status. For Yenisei peoples frame-shaped JHs represent the "voice" of a local deity in charge of successful hunting (126), whereas metallic bow-shaped JHs are an attribute of power and prestige, entitling the owner to protection (131). The ideological divide also involves an aesthetic aspect. Across the entire southeastern end of Russia, all cultures that contain both frame-shaped and bow-shaped JHs consider the former suitable for learning the JH, but not for "serious" music-making (Sheikin, 2002, p. 131)[70].

Hence, each model follows its own course. Around Kama, bow-shaped JHs supplanted frame-shaped JHs (Yakovlev, 2001). Generally, the archaic polytheistic belief in sacred places yielded to a monotheistic cosmology under the influence of Christianity after Russian colonization (Alekseyev, 1992, p. 215). This weakened the kin-tree/kin-bone correspondences that provided ideological support to the frame-shaped tradition. The strengthening of top-deity cults, the Christian-like dichotomy of good/evil, and divine protection supported the "ownership-based" (not kin-based) protection of bow-shaped JHs.

The autonomy of frame- and bow-shaped traditions is corroborated by the late arrival of metallurgy to the north Pacific coast (**Figure 8**). The Chinese ideograph for the metallic JH (*tieyehuang*) confirms that the bamboo JH (*huang*) preceded it (Tadagawa, 1991).

The scenario in which the bow shape was the descendant of the frame shape also received organological support (Sachs, 1917; Dournon-Taurelle, 1975; Sheikin, 2002; Suzukei, 2010). Organological analysis of European archaeological finds indicates that European bow-shaped JHs form their own lineage of morphological development from earlier Asiatic samples brought to Europe through trade (Kolltveit, 2006). Currently, the archaeological consensus holds the organic

[65] In Taiwan and Indonesia there are very few examples of indigenous musical traditions that use bamboo frame-shaped JHs with metallic lamella (Kolltveit, 2016). This is also the case in Vietnam (Wright, 2001). However, judging by available recordings of such instruments, this hybrid construction does not substantially differ in spectral texture from the frame-shaped instruments that are made just out of bamboo. On the territory of Russia, bow-shaped JHs are always metallic (Sheikin, 2002, p. 129).

[66] There are some vague ethnographical references indicating that wooden JHs have been used in Flanders, Hungary and Ireland—however, such cases seem to be exceptional, leaving no archaeological or historic traces (Kolltveit, 2009). It is likely that wooden instruments such as Slovenian *drumlja* present a relatively modern development, called to reduce costs and thereby increase sales of JH in modern market conditions.

[67] Low quality of mass produced instruments is known to negatively affect the standards of JH music produced on imported JHs—as compared to hand-crafted instruments of indigenous JH traditions (Morgan, 2017). Metallic instruments, less responsive to the player's technique, must have further limited the spectral content of JH texture, increasing the dominance of a single melodic "part" even more.

[68] Such scenario has been confirmed in relation to Yakutia where organic frame-shaped JHs were found to precede the metallic bow-shaped JHs (Dyakonova, 2017). The same applies to the peoples of Volga-Urals (Yakovlev, 2001), Kama (Golubkova and Ivanov, 1997), Sakhalin (Mamcheva, 2012).

[69] Thus, amongst Khanty and Mansi, frame-shaped *tumran* is strictly female, related to wedding rituals of fortune-telling and "avoiding" men (Sheikin, 2002, p. 127). Across Southeastern Siberia bow-shaped JHs are used exclusively by men in funeral lamentations, whereas females use bowed pikolute instead (131).

[70] An interesting case presents the tradition of Ude JH. Despite the fact that Ude hunters consider frame-shaped JHs as being children instruments, nevertheless, they report that hunters must play it after setting their hunting traps to ensure that the prey will be trapped (132). This testifies to a more ancient origin of the frame-shaped construction in hunter-gatherer's societies—in contrast to the bow-shaped construction that did not exist before the Metal Age. The same applies to their respective musical textures.

**TABLE 2 |** The correspondences between the complexity of manufacturing JHs, JH's spectral texture, and JH's sound qualities.

| # | Material | Availability | Manufacturing | Texture | Complexity | Sound |
|---|---|---|---|---|---|---|
| 1 | Grass (Leymus) | Readily available in grasslands, albeit seasonal, and disposable | Needs no tools, takes about a minute to make for anyone (Mamcheva, 2012)—minimal expertise required | "Scattered" (haphazard) polyphony of 5 melodic parts | All parts are indistinct and melodically similar, most likely only the dynamically marked tenor is controlled, while the other parts are "aleatoric," all parts are hemitonic or chromatic, including bass | "Hollow" clicking, very short, dry, yet soft and muffled, clattering at high intensities |
| 2 | Bamboo (Bambusa, Phyllo-stachys, Dendro-calamus) | Readily available in a warm and moist climate, evergreen, non-disposable | Needs basic cutting tools, takes days to dry and minutes to make with stone tools (Bar-Yosef et al., 2012)—manufacturing is somewhat complicated by anisotropy—some expertise required | Duplum polyphony of 4 parts: upper pair carries 2 melodies, lower pair accompanies | All parts are clearly defined: alto has the principal diatonic melody, soprano—secondary chromatic melody, tenor—a dyadic or triadic ostinato figure (or a chord), bass—a pedal monotone; alternatively, tenor can carry the 2nd melody while soprano—the pedal chord with slight variations | "Hollow" rattling or rasping, relatively dark, rich, and quite homogenous tone, prolonged decay, introducing tremolo at high intensities |
| 3 | Wood (Larix, Betula, Salix, Alnus, Picea, Cedrus, Populus) | Available in accessible lands with temperate mesothermal or continental microthermal climate, all year round, non-disposable | Needs cutting and shaving tools, takes days to dry and about a few hours to make with stone tools (Crabtree and Davis, 1968)—manufacturing is greatly complicated by anisotropy, orthotropy, hygroscopy, variance in quality, knots and spiral grains—extensive expertise needed | Triplum polyphony of 5 parts: middle trio carries 3 melodies, marginal parts accompany (the upper part can be absent) | Melodic "trio" is partially differentiated between secondary anhemitonic melody in tenor and primary chromatic in alto, but not in regards to the tertiary chromatic melody in soprano, while 2 pedal parts—monotone in bass and chordal in descant—frame the melodic "trio" (alternatively, there could be just 4 lower parts—without a descant) | Clattering, dense and rather dry, with a distinct "clang" on the onset, which is more pronounced at high intensities, overall bright tone with medium long decay—great variety in tone for different woods |
| 4 | Bone (cortical) and antler | Available from animals practically everywhere, all year round, non-disposable | Needs cutting and scraping tools, takes weeks to soak (Robinson, 1942), hours to cut and about a day to carve with stone tools (Sidéra, 2011)—manufacturing is much complicated by anisotropy, hygroscopy—extensive expertise needed | Triplum polyphony of 4 parts: upper trio carries 3 melodies, while bass—pedal chord or dyad | Melodic "trio" is poorly differentiated except for the bass that sustains a harmonic pedal (triad or dyad), tenor usually carries the principal hemitonic melody, while alto—the secondary chromatic melody (but this can be reversed), and soprano carries the tertiary melody—chromatic like secondary melody | "Hollow" clucking, less definite in pitch than in bamboo and wood, more gentle, well-rolled decay, quite homogenous, medium bright tone, quite long-lasting |
| 5 | Metal (copper and its alloys, iron, steel) | Available in hard to access ore deposits, usually in mountains, nearly all year round, the most long-lasting | Needs finding of ore, mining and transporting it to the smelting site, pit-digging, carbon and hermitization to sustain high heat, stationary equipment to extract, smelt and forge metal, takes hours to heat (Coghlan, 1939), days to forge (Bronson, 1996) and tune (Jaago, 2009)—demands professional expertise | Homophony of 3 parts (unavailable on non-metallic JHs)—with the melody in the middle, but possibly in the upper part while chords in tenor (4- and 5-part textures, like those on bamboo or wood are also possible, especially on idioglot JH) | All parts are perfectly differentiated and easy to track: the pedal bass, the rich melody in tenor, and the ostinato "chords" in alto (the upper pair can be reversed); in alternative scenarios an extra part can be added in soprano, which contains a pedal chord, while tenor or alto carries a secondary melody—overall, metallic JHs produce the greatest variety of textures; the principal melody usually has the most degrees of all materials (especially for steel), hemitonic or chromatic, vs. monotone or dyadic bass and triadic-based chord in tenor or chromatic (cluster-like) chord in soprano | Ringing or buzzing, resonant, very full and rich, the brightest, the most well-defined in pitch and the most homogenous in tone of all JH materials, with the longest decay—great variety in tone for different metals and alloys (metal seems to support much greater control of fine tonal detail than non-metallic materials) |

*The numbering in the table reflects: the progressive increase in complexity of the manufacturing technology, the decrease in availability of the material to the JH manufacturer, the increase in discretization and variability of the components of the spectral texture, and the increase in timbral richness, fullness and homogeneity of the sound—that characterize each of the materials [for a more complete information see* ***Table 1*** *in the Appendix and "Textural typology for Appendix.xml" (****Supplementary Material****)]. The identified pattern of evolution could potentially apply not only to JH, but to other musical instruments made of the same materials. This possibility calls for additional research.*

**TABLE 3 |** Classification of JHs based on similarities and differences in sound quality, spectral texture types and playing techniques between the JHs made of different materials.

| # | Raw material for making | Traditional JH models of this type | Their standard organological description | Basic playing motion of a hand | Characteristic sound quality | Textural typology |
|---|---|---|---|---|---|---|
| 1 | Grass leaves, tree bark or tree chips (or splinters), or twig of a shrub or a bush | Nivkh and Ulch *konga chnyr* (Mamcheva, 2005), Yakut *mas khomus* (Dyakonova, 2017), Tuvan *charty-komus* (Suzukei, 1989), *yiash-komus* and Telengit *taia-komus* (Suzukei, 2006) | Idioglot plucking idiophone, frameless lamella (frame-shaped, but open on one side or wishbone shaped)—classified as "proto-JH" by Sheikin (2002) | Touching the leaf, gently tapping on the chip with a finger, or pinching it with two fingers, tapping on the twig | Non-percussive very soft hollow clicking sound, homogenous and simple (the spectral components are poorly isolated from each other, no way of playing melodic modes based on harmonic series, no sustained drone component)—brighter and denser | "Maximal" polyphony of 4–5 indiscrete poorly defined parts, each of which is moving at the same time and sounds similar to the other ones |
| 2 | Bamboo culm, wood beam or fragment of fresh cortical bone | Nivkh *tyf-kanga* and *khar-kongon* (Mamcheva, 2005), Ainu *mukkuri* (Nobuhiko, 2008), Evenk *kongiipkavun* (Sheikin, 1986a), *Tumran* of the Ugric peoples (Alekseyenko, 1988) | Idioglot plucking idiophone, framed JH whose plate contains a rope for energetic plucking or pulling as well as a handle (the lamella is framed from all sides, is elongated and bottle-shaped) | Abrupt pulling of a rope by the right hand—JH supports a greater diversity of technical devices (due to greater intensity that exposes more detail) than for the proto-JHs | Loud, low, dry, and harsh sound that lacks in homogeneity ("dirty") due to either rattling, rasping or bifurcated "clang" and good or superb separation of parts (can stress few parts at the same time, plays the pentatonic harmonic scale and uses the clear bass drone) | "Optimized" polyphony of 4–5 discrete parts, most of which are functionally and/or thematically different (melody, counter-melody, accompanying melodic figure, or pedal chord, dyad or tone) |
| 3 | Metal ores | Nivkh *pasla vyt' kanga, zakanga* and *vychranga* (Mamcheva, 2005) and Buryat *aman khuur* made of local "red copper" (Simukhin, 2006) | Idioglot plucking idiophone, bow-shaped JH whose lamella is bent into a hook and attached to the frame (shaped like a horse-shoe); an alternative construction is heteroglot, which most commonly utilizes brass and reproduces the bamboo/wood constructions and the corresponding polyphonic textures | Hitting or strumming the lamella forward, backward, pressing or just holding the lamella with a finger, firm or loose holding of the ring by teeth or lips, right or left shift of the frame (Sheikin, 1986b)—greater diversity than in organic JHs | Prolonged ringing or buzzing sound, harmonically richer and louder than JHs made from organic materials, especially in high range, even greater resolution of detail (displays the content of each part, can play chromatic and microchromatic melodies, place the melody in any part except in the bass, support superb isolation of spectral components) | "Simple" or "complex" (with a few different accompanying parts) homophony with a single melody that is more diverse (use more degrees) than polyphonic melodies of non-metallic JHs—for idioglot JHs; yet heteroglot metallic JHs in addition to homophonic texture use the same polyphonic textures as bamboo/wood JHs |

*In general, different hand action in production of sound on a JH corresponds to a different sound quality (Mamcheva, 2012, p. 53–55) and textural typology (see the Appendix and listen to the audio examples in the archive "Audio for Appendix" in* ***Supplementary Material****). Their correspondence is somewhat complicated by the cumulative nature of indigenous JH traditions, exemplified in Nivkh JHs (Mamcheva, 2005): newer technologies of manufacturing JH tend to adopt the standards of playing from older previously existing technologies, before the novel constructions allow to forge new standards (see Appendix in* ***Supplementary Material*** *for the discussion of this issue). Therefore, metallic JHs are used to generate textures characteristic for bamboo, wood or bone JHs—as well as proprietary "metallic" textures, unproducible on JHs made of organic materials.*

JHs as the prototype for the metallic via westward expansion from northeast Asia (Fox, 1988; Wright, 2004; Kolltveit, 2006; Honeychurch, 2015; Aleksandrova, 2017; Turbat, 2017; Oleszczak et al., 2018). Another expansion is likely to have followed from South China to Austronesia (Blench, 2004).

The time frame for the genesis of organic JHs can be established by dating the migration of the Siberian population to America. Many North Amerindians use the PS (Ojamaa, 2002).

**Example-56**

The topahti—a personal Nootka song of inherited origin, performed by Joe Titian. This topahti was given at the inter-tribal marriage between Nootka and Kwakiutl as a dowry, and permitted for performance only by its owner and her children (Halpern, 1974) (http://chirb.it/NvahDq).

But no JH usage is known before the Western colonization (Wright, 2011)[71]. This is hard to reconcile with the evidence for the common genetic ancestry of Native Americans and

[71]JHs, rather popular amongst modern Inuits (Nattiez, 1976), have been introduced in Alaska, Canada and Greenland from Western Europe (Whitridge, 2015). The JH-like use of the feather of an arctic bird, common eider (Somateria mollissima), was observed amongst the Inuits of the Belcher Islands and northern Quebec (Oakes, 1991). They hold the feather in their mouths and beat it with one hand. A photograph of this is published in the LP "Inuit Throat And Harp Songs" (Green and Hodge, 1980). Most likely, feather is used as a more affordable material to substitute the imported metallic JH. However, the possibility that the feather use is a local invention that might have preceded JH trade cannot be completely discarded. However, even in this case, such feather use remains totally isolated.

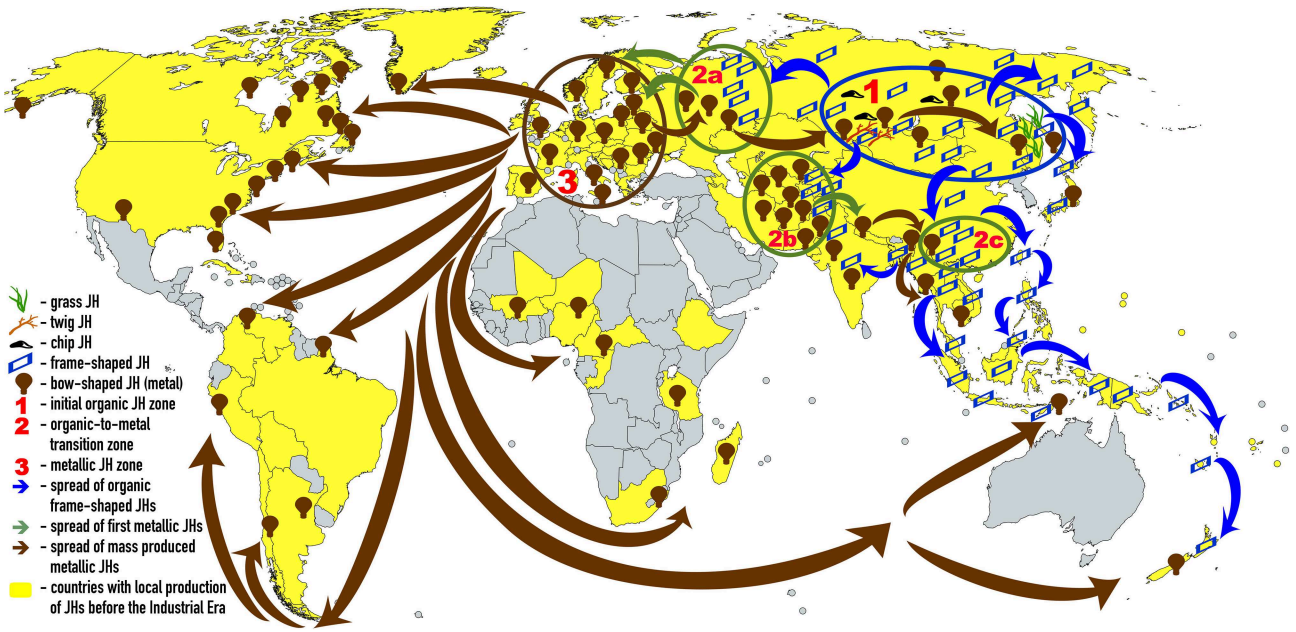



**FIGURE 7 |** World distribution of JH made of organic and metallic materials prior to the Industrial Period. This map combines archaeological and ethnographic data on local production and consumption of JH across the world, excluding mass international trade from Europe, in order to establish the routes of the spread of metallic and non-metallic JH indigenous traditions. Yellow color marks those countries where JHs were produced and widely consumed by local population. Such countries were selected based on the integration of findings of linguistic research on the presence of native words for JH in local languages (Fox, 1994; Sheikin, 2002, p. 411–503; Bakx and Crane, 2017), of archaeological finds (de Ramón and Rivera, 1982; Beck et al., 1983; Barr, 1994; Kungurov, 1994; Yakovlev, 2001; Pignocchi, 2002, 2004; Crane, 2007; Wright and Impey, 2007; Wright, 2011; Honeychurch, 2015, p. 20; Whitridge, 2015) [plus references from **Table 1**], and of geo-ethnographic research (Vertkov et al., 1963; Galaiskaya, 1973; Fox, 1988, 1994; Sheikin, 2002; Wright, 2004; Yesipova et al., 2008; Suzukei, 2010; Mamcheva, 2012). Geo-ethnographic data was also used to distinguish between the areas of distribution of bow-shaped metallic and frame-shaped organic JHs. Five icons with corresponding pictures mark the geographic locations of musical cultures that use JHs made of different materials. Red numbers indicate the zones of distribution for different types of JHs—encircled in different colors. (Number 1) Marks the blue oval that shows the area where organic-made JHs have been prevalent (see **Figure 4**). This area most likely constitutes the cradle of JH music. (Number 2) Marks the green ovals that encircle the areas where metallic JHs have coexisted with organic ones. It is in these areas, that presumably, the transition from organic to metallic instruments took place. Two areas [(2a) and (2b)] correspond to the oldest centers of metallurgy (see **Figure 8**). The transition probably occurred first in 2a), where the access for import of organic JHs from zone (1) was geographically broader and easier than in (2b), and organic JHs were likely to have been established prior to the wide spread of metallic artifacts. Transition in (2c) area most likely occurred much later because of the later arrival of metallurgy (see **Figure 8**). (Number 3) Marks the brown oval that encircles the area where exclusively metallic JHs were cultivated. The color-coded arrows show the probable direction for spreading of the organic and metallic JH technologies. Green arrows indicate the distribution of the initial handicraft metallic JHs that might have emulated popular wooden/bone/bamboo types. Brown arrows indicate the distribution of the mass-produced bow-shaped metallic construction, prevalent in Europe. Across the ocean, these arrows connect the European dominions with their offshore colonies. Over the land, brown arrows indicate the chain of cultural contacts that led to the establishment of local production centers of metallic JHs. Evidently, one of the routes of this spread went contrary to the original spread of the organic JHs—from European West to Far East, over the Steppe Belt's borderline with the northern forests via the chain of neighboring Turkic cultures that all favor metallic JHs. This route earned the nickname of "Fur Route" (Rubinson, 1992; Bunker, 1993). Another route went from Central Asia to China via the Silk Road—historic as well as prehistoric, as recent research suggested (Christian, 2000; Kuzmina, 2008; Barinova, 2013).

south Altaians 23000–18000 BCE (Schurr, 2015). How could an instrument so important for Siberian peoples be totally missing from their American descendants? The Altai-Sayan region is home to the JH. According to genetic and linguistic evidence, Yakuts, whose national symbol is the *khomus*, descend from there (Pakendorf, 2007). Siberia and Alaska remained connected until 8000 BCE, but glaciers blocked Alaska from the rest of the continent until 11000 BCE (Dixon, 2015). Sequencing mitochondrial genomes from pre-Columbian South American skeletons 8,600–500 BP indicates that a small population entered the Americas through a costal route 16,000 BP (Reich et al., 2016). This date comes close to the earliest archaeological evidence of a human presence on the North American continent: 13,000 BP (Anderson and Bissett, 2015).

The most plausible scenario is that the JH tradition did not exist along the northeast Asian coast before 10,000 BP, when access to America was open. Throughout the Quaternary Northeastern Asia enjoyed a remarkably stable climate, and lowlands remained ice-free (Zamoruyev, 2004). The prevailing landscape of the non-glaciated area of Altai was desert-steppe, covered with grass, with patches of woody vegetation at stream valleys—changing into coniferous forest or forest-steppe only in the Holocene (Hais et al., 2015). Similar vegetation covered Mongolia and Tuva. Modern forest-steppe regions, e.g., Ob'-Irtysh Baraba, remained steppe throughout the early Holocene until 5500 BP (Zhilich et al., 2017). In highlands such as the Chuya Alps, the first forestation is dated 7000 BP (Agatova et al., 2012). Forestless landscape stretched toward North

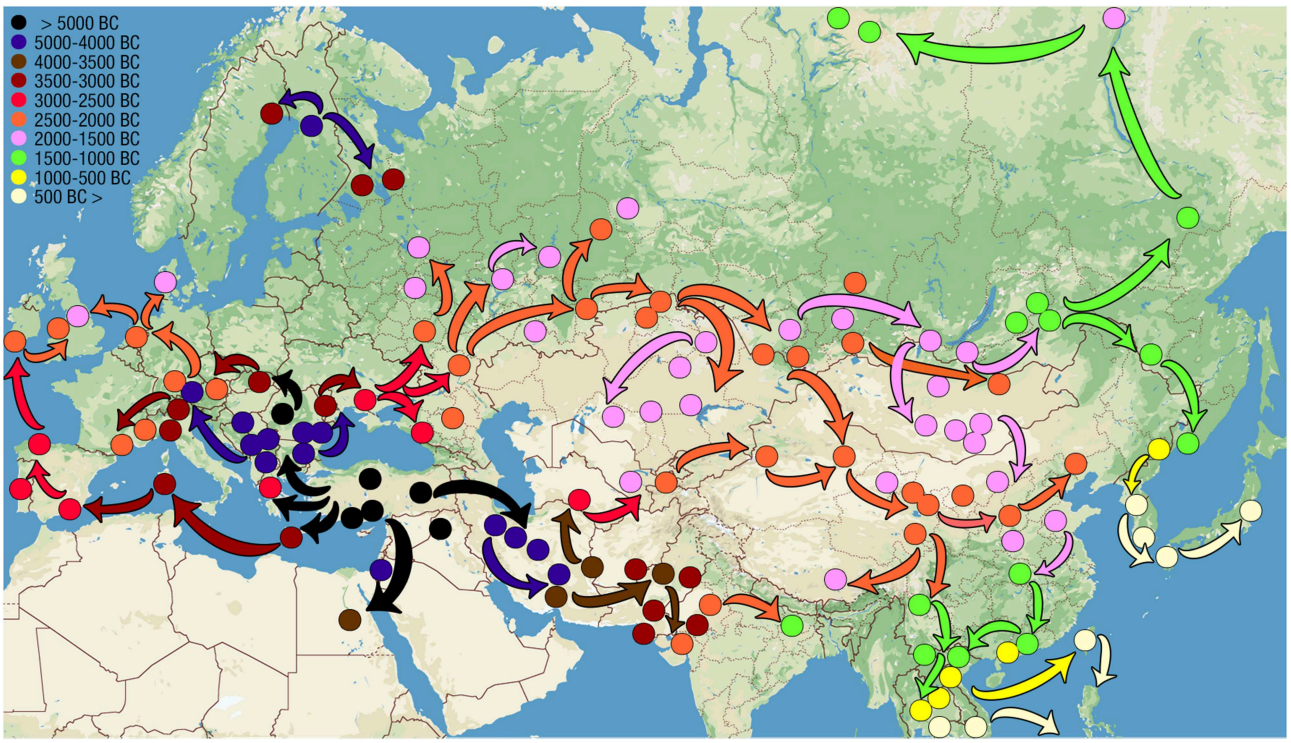



**FIGURE 8 |** Historic spread of copper metallurgy in Eurasia. This map displays the locations of earliest regional centers of smelting copper ores—according to the available archaeological research on the earliest metallurgy (Chernykh, 1966, 1992, 2012; Sunchugashev, 1975; Zhuravlev, 1977; Sergeyeva, 1981; Prakash and Tripathi, 1986; Rybakov, 1987; Kon'kova, 1989; Thiel, 1989; Mishra, 1994; Reedy, 1997; Kiriushin, 2002; Tylecote, 2002; Nguyen, 1986; Baipakov and Taimagambetov, 2006; Simukhin, 2006; Yanin, 2006; Ciarla, 2007; Hauptmann, 2007; Kaniuth, 2007; Park and Gordon, 2007; Subbotina, 2008; Herva et al., 2009; Roberts et al., 2009; White and Hamilton, 2009; Radivojević et al., 2010; Erb-Satullo, 2011; Higham et al., 2011; Wan, 2011; Potts, 2012; O'Brien, 2014; Garner, 2015; Gelegdorj, 2015; Mei et al., 2015; Hung and Chao, 2016; Huo, 2016; Tripathi, 2018). The routes and timeline of its spread suggests the spread of metallic JHs along with the new homophonic tradition of JH music. The timeline is indicated by color-coding of the circle-icons that mark the location of smelting sites: the darker—the older. The epicenter of metallurgy is in modern Turkey and Iraq, although a concurrent independent center might have existed in Vinca culture, in Balkans, Belovode (Radivojević et al., 2010). Anther independent center emerged in Fennoscandia (Herva et al., 2009) and the neighboring Karelia (Zhuravlev, 1977). Copper metallurgy spread from Balkans to Ukraine, thereon bifurcating, with one branch heading to the northeast forest zones of Russia (up to Ural Mountains), while the other moving southeast through the Steppe Belt all the way to Zabaikalye and Mongolia. There, again the route bifurcated: north, going to Yakutia, and south to Primorye. The other wave from the epicenter reached Iran, branching in 3 directions: toward Indus' valleys, Himalayas, and Middle Asia, through Turkmenistan, Uzbekistan and Kazakhstan to the Uyghur territories via the route that later became known as the Silk Road (Kuzmina, 2008). Around Lop Nor this branch met the outshoot of the Steppe Belt route and proceeded through modern China to its North and South kingdoms. The southern branch eventually hit Indochina, then heading to Indonesia and Pacific islands. This entire area, Manchuria and Japan were the last to develop metallurgy—no earlier than in Middle Ages. Therefore, if for Volga-Ural area frame-shaped JHs co-evolved with the metallic JHs, for cultures at the eastern end of Eurasia bow-shaped JHs came from West and formed its own special niche, different from the already established frame-shaped JH cultures. The bow-shaped tradition must have taken a long time to develop—as metallic artifacts were becoming more affordable to local population. Another zone where frame- and bow-shaped JHs form different cultural traditions is the vast area from southeastern Indochina to Papua New Guinea. The two zones where metallic bow-shaped JHs must have penetrated concurrently with organic frame-shaped JHs where both might have shared the same or similar forms of use, are Himalayas-Pamir-Tian-Shan mountains and Yunnan-Myanmar area.

America, including Beringia (Hoffecker and Elias, 2007). The tundra belt extended to 57°N, then turned into steppe that covered most of Eurasia (Tarasov et al., 2000). Forestation was prominent only much farther south, at Taiwan's latitude, during the late Pleniglacial (Hope et al., 2004)—too distant from Beringia.

The scarcity of trees would have prevented the formation of the "person/kin/ethnos=tree/species/forest paradigm" that underlies the Siberian timbral music culture. The emergence of "sacred" tree cults depends on trees' importance in sustaining a human population, which requires that they be available in abundance. Otherwise, plants' proprietary "voices" cannot be discovered by population groups large enough to institute a musical tradition.

## CONCLUSION

Not all known music systems abide by the principle of discrete changes in pitch. In fact, many music cultures rely on timbral changes in their TO. One trait such cultures share is the prevalence of the "personal" over the "collective" use of music, upholding the opposition between "timbre-centered" (definite timbre/indefinite pitch) and "frequency-centered" (definite

pitch/indefinite timbre) music systems. Current psychoacoustic research seems to support this opposition.

The large concentration of "timbral" cultures in the former USSR prompted research (often emic[72]) on their TO, promoted by the centralized infrastructure of research institutions, "top-down" funding (official ideology favored "people's music"), and the intense development of ear-training research. The accumulation of vast ethnographic and archaeological data over much of the twentieth century made it possible to create a "historic" perspective on the evolution of folk music. The integration of sciences within Soviet academe supported a multi-disciplinary approach that combined fieldwork with experimental testing. The gathered evidence suggests that "frequency-" and "timbre-centered" musics contrast each other in their TO, patterns of usage, and their social implementation.

Timbral TO relies on PS in vocal, and JH musicking in instrumental domains. Every member of such traditional society possesses at least one PS that serves for personal identification. Musical elements usually indicate family relations and birthplace through similarities with other individuals' PSs. Locals have a keen ear and memory for cross-relating thousands of PSs throughout their life (better than for remembering faces)[73].

PS contributes to mental health in the traditional Siberian lifestyle. Ongoing singing throughout one's day secures consistency in self-consciousness under environmental stress, common in severe Siberian conditions, thus protecting one from psychological disorders. Pathological singing accompanies multiple personality disorder (*meneryia*) and sleep-singing disorder (*tyyl yryata*), most probably related to the dysfunction of the PS circuit.

In traditional societies, PS lowers the standard of musicality to maximize singing accessibility. Close relatives sing their PS *for* aphoniacs. An indefinite intervallic structure and an absence of the notion of "wrong notes" make a PS performable by almost anyone. Playing JH is even more accessible.

The JH opposes PS by "de-personalizing" one's natural voice via the JH's vocoding mechanism. While camouflaging one's voice, the JH "impersonates" a wealth of environmental sounds. Its exceptional mimicry allows for re-creating the "real world" musically through the timbral abstraction of sounds and their arrangement according to principles of phonetic symbolism.

The JH addresses the objective aspects of reality, PS the subjective. Together, they form a coherent musical worldview, which explains why numerous North Asiatic ethnicities have so few "pitched" musical instruments: They simply did not have much need for them.

Indefinite-in-pitch rhythmo-timbral "themes" effectively represent:

- for PS—one's daily chores ("person"=monotheme, "chores"=variables in PS's variations);
- for JH—one's environment ("person"=low drone, "environment"=mid/high spectrum).

Both rest on the traditional hierarchic paradigm: "person"-"kin"-"ethnos"="tree"-"tree-species"-"forest." Such a paradigm must have formed in southeastern Siberia/Manchuria ≈10,000 BP. JHs made of ancestral plants were initially used as talismans. The accumulation of onomatopoeic devices and conventions of phonetic symbolism created "timbral words," "timbral phrases," instrumental patterns emulating PSs, and, eventually, proprietary JH TO, where the key role belongs to a set of "harmonic templates." Each template carries a specific semantic value, enabling linguistic-like semiosis: meaningful elements comprise meaningful components, but convey emotional rather than referential information. This development must have directed the entire evolution of Eastern Eurasian timbre-based music systems in opposition to the frequency-based music of the Eurasian West[74]. *The JH was for the prehistoric East what the bone pipe was for the prehistoric West.*

The personal nature of PS/JH traditions stems from their reliance on timbre, which is fundamentally unsuitable for collective musicking. Collective performance of PSs or JHs would make their message unrecognizable. Only frequency-based music allows the tradition of collective performance to form and continue.

Western frequency tradition, exemplified by Aurignacian pipes (Morley, 2013), might have African roots. Africa underwent two major demographic expansions prior to the Aurignacian, enabled by its tropical ecology (Lahr and Foley, 2001). The second expansion (86,000–61,000 BP) carried haplogroup L3 outside of Africa (Atkinson et al., 2009). Genetic evidence suggests that non-Africans descend primarily from this migration, whose maximum falls on 70,000 BP, coinciding with the improved climate in East/Central Africa (Soares et al., 2012). At that time, the East African effective population size was at least 10,000 people (Relethford, 1998), vs. the census maximum of 3,700 in Gravettian Europe[75].

Tropical environments generally support greater population densities than those at higher latitudes (Layton and O'Hara, 2012). Environmental conditions impact demographic density

[72]Thus, three of the authors of this paper were raised in the native cultural environment of Yakutia, and therefore combine a first-hand "emic" knowledge about its indigenous music (and language) with an expertise in musicological analysis of such music.

[73]Long-term memory for melodic structures is known for remarkable longevity and capacity (Halpern and Bartlett, 2010). Accurate memorization of music can be life-long (Bartlett and Snelus, 1980). Music seems to provide superior encoding, evident in the phenomenon of earworms, unique to the domain of music (Halpern and Bartlett, 2011).

[74]For more information on the role of flute-like instruments in determining a music system see "Tonal organization in tuning of Paleolithic and Neolithic pipes" (Nikolsky, 2015, Appendix-2).

[75]The median ratio of effective population size to census size remains exceedingly low across 43 mammalian species, constituting on average 0.003 (Nei and Graur, 1984). The method of mismatch analysis, introduced by Rogers and Harpending (1992), allows for estimation of size and timing of ancient demographic based on the genetic data. Population growth at the time of expansion is estimated to be 100-fold or even greater (Rogers, 1997). And the timing of it clusters around 50,000 BP (Sherry et al., 1994). The African expansion of this time suggests a demographic picture with sparse population outside of Africa, while a rapid population explosion inside of Africa (Relethford, 1998).

the most, and even the tropical desert supports a denser population than the polar biome (Tallavaara et al., 2018). Intergroup connectedness also drops at higher latitudes; the large effective population of connected groups in the African Middle Stone Age contrasts with stochastic variation without linear trajectories in the contemporaneous European Mediterranean region (Malinsky-Buller and Hovers, 2019). Sparse groups' migration leads to frequent losses of gained cultural skills. Steady post-Gravettian demographic growth triggered the cumulative cultural complexity that characterizes behavioral modernity (Shennan, 2001). Part of this was the consistent increase of foragers' group size (Grove, 2012). By 45,000 BP, the median effective population size in Europe equaled that which sub-Saharan Africa had reached 101,000 BP—alongside the markers of modern behavior (Powell et al., 2009).

Population growth promotes group cohesion, territoriality, ethnogenesis, and language formation (Robb, 1993). "Frequency music" likely followed suit. Steady demographic growth accompanied the Neolithic "revolution" and civilizations' rise (Hassan, 1981)—together with "frequency-based" music (Nikolsky, 2016). *Collective music-making was part of the "demographic expansion package" designed to consolidate and empower the "tribe" to grab and hold its territory*.

Radically different is the demography of northeastern Eurasia. Siberia is famous for its immense land (13,100,000 $km^2$), sparse population (200,000 before Russian colonization), and the so-called everlasting importance of hunting/gathering for sustenance (Naumov, 2006). Such population density—0.065 person/$km^2$–approximates the 0.036 person/$km^2$ maximum of Magdalenian Europe (Maier, 2017). Harsh living in the traditional lifestyle makes landholding not a viable strategy. Constant migration by small "packs" requires marking and regulating the sharing of territory between all neighbors, helping each other to survive (Funk, 2006). Therefore, local beliefs assign power not to the "tribe" but to the spirit-masters of landmarks on whose disposition human "tenants" must rely. This has laid the foundation for the PS/JH pan-Siberian framework. And since the climate in Northeastern Asia remained remarkably stable throughout the Quaternary (Zamoruyev, 2004), it is reasonable to believe that the institution of PS/JH characterizes the music culture of local prehistoric people.

It can hardly be a coincidence that the area where JH remains to be the principal musical instrument in scarce instrumentarium is identical with the area of the greatest concentration of Denisovan genomes. The highest levels of Denisovan ancestry is found in Oceanic populations (Vernot et al., 2016). Denisovan genomes are also present in Eastern Eurasians and Native Americans (Qin and Stoneking, 2015). Denisovans may have interbred with early humans over the territory of Northern China (Martinón-Torres et al., 2017), where the oldest JHs were unearthed. Longevity of the JH dominance might constitute a distant remnant of the Denisovan timbral music tradition, preserved in the refugium of isolated Pacific islands, north-Chinese deserts and Altai mountains. Neanderthal heritage could entail the "frequency music," carried from Africa by Homo Heidelbergensis. The latter adapted to the northern latitudes as opposed to southern Homo erectus—and so is the case with Neanderthals as opposed to Homo Heidelbergensis (Grove et al., 2012). Either of them might have adapted the southern collective "frequency music" to the northern ecosystems, generating a new[76] personal "timbral music."

The sparse Neanderthal (French, 2016) and Denisovan (Meyer et al., 2012) populations of the Pleistocene Altai (Buzhilova et al., 2017) might have also subscribed to "timbral music". Homo's "timbral music" either descended from Neanderthals and Denisovans, or "downgraded" from the European "frequency music" carried by "Ancient North Siberians" from the West ≈38,000 BP (Sikora et al., 2019).

## AUTHOR CONTRIBUTIONS

EA and AN conceived the presented idea. VD organized the recording of Jaw Harps at the Jaw Harp Museum in Yakutsk, collected all the necessary ethnographic materials and provided the most recent data of the fieldwork research in Siberia and Russian Far East. She verified the information on traditional musical instruments in Siberia. IA performed on different Jaw Harps and had them recorded and acted as a consultant in relation to the indigenous Siberian traditional music and matters of prosody and phonology of Jaw Harp playing and singing. EA provided his archive of the recordings and supervised the application of his methodology of the analysis of tonal organization of indigenous folk music. AN conducted the acoustic and musicological analysis of the provided material, conceived the method of analysis of spectral textures, corroborated all the findings with the research in the former USSR, modern Russia, and Western countries, wrote the manuscript and created the figures and tables for it. EA edited the manuscript. AN translated it into English.

## ACKNOWLEDGMENTS

We are grateful to Valentina Suzukei, Natalia Mamcheva, Oksana Dobzhanskaya, Yurii Sheikin, Tatyana Ignatyeva, Gjermund Kolltveit, Erkin Alekseyev, Ricardo Eichmann, Leon Crickmore, Jeremy Montagu, Triinu Ojamaa, Jaan Ross, Iraida Seliutina, Nikolai Urtegeshev, Tatiana Ryzhykova, Albina Dobrinina, Tatiana Daineko, Josef Jordania, Alexandra Khaltanova, Yevdokiya Sergina, Spiridon Shishigin, Alevtina Everstova, and Vyatcheslav Shchurov for providing us with the necessary materials for this project. We appreciate the thorough critical input from Theodor Levin that was instrumental for shaping the scope of our research. Special thanks to Erkin Alekseyev, Nastya Petrova, and Aksenty Beskrovny for their recordings of Jaw Harp, and to Sheila Bazleh for editing the text of this manuscript. Also we would like to thank Brian C. J. Moore, Viktor Reznik, Stephen McAdams, Pantelis N. Vassilakis, Steven Brown, Jaan Ross, and Frank Scherbaum for their help in elaborating the

[76]If the Divje Babe pipe is indeed not a byproduct of the hyena bite (Diedrich, 2015), but rather a Neanderthal musical instrument (Turk, 2014), then the Neanderthal expansion to Altai would have created a Paleolithic border between the frequency- and timbre-based music.

optimal approach to taking acoustic measurements of Jaw Harp recordings and interpreting their results.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.03051/full#supplementary-material

**Data Sheet 1 |** Appendix. Spectral texture.

**Data Sheet 2 |** List of audio examples.

**Data Sheet 3 |** Textural typology for Appendix.

**Presentation 1 |** Audio for the article.

**Presentation 2 |** Audio for Appendix.

**Presentation 3 |** Figures for Appendix.

## REFERENCES


Aberle, D. F. (1952). Arctic hysteria and latah in Mongolia. *Trans. N. Y. Acad. Sci.* 14, 291–297. doi: 10.1111/j.2164-0947.1952.tb01125.x

Adkins, C. J. (1974). Investigation of the sound-producing mechanism of the jew's harp. *J. Acous. Soc. Am.* 55, 667–670. doi: 10.1121/1.1914580

Agapitov, N. N., and Khangalov, M. N. (1885). Data for research on shamanism in Siberia. Shamanism amongst Buryats from the Irkutsk Province [Материалы для изучения шаманства в Сибири. Шаманство у бурят Иркутской Губернии]. *Proc. East Siberian Imperial Russ. Geogr. Soc.* 14, 1–61.

Agatova, A. R., Nazarov, A. N., Nepop, R. K., and Rodnight, H. (2012). Holocene glacier fluctuations and climate changes in the southeastern part of the Russian Altai (South Siberia) based on a radiocarbon chronology. *Q. Sci. Rev.* 43, 74–93. doi: 10.1016/j.quascirev.2012.04.012

Agrawal, P. K. (2018). *Indian Culture, Art and Heritage*. New Delhi: Prabhat Prakashan.

Aizenshtadt, A. M. (1982). "With kets and selkups [У кетов и селькупов]," in *Music of Siberia and Far East [Музыка Сибири и Дальнего Востока]*, ed I. Romashuk (Moscow: Sovetskii Kompozitor [Советский композитор]), 175–204.

Aleksandrova, N. (2017). "Frame-shaped Jew's harps in archeological monuments at the Kama and Vyatka interfluve [Пластинчатые варганы в археологических памятниках междуречья Камы и Вятки]," in *Problems of Organology*, ed I. V. Matsiyevskii (Sankt-Petersburg: Russian Institute of History of Arts [Российский институт истории искусств]), 83–91.

Alekseev, I. Y. (1988). *The Art of Playing Yakut Khomus [Искусство игры на якутском хомусе]*. Yakutsk: Preprint.

Alekseyenko, A. M. (1988). "Musical instruments of the peoples of North East Siberia [Музыкальные инструменты народов севера Западной Сибири]," in *Material and Spiritual Cultures of Siberian Peoples [Материальная и духовная культура народов Сибири]*, ed C. M. Taksami (Leningrad: Nauka, The Institute of Ethnography of the USSR Academy of Science), 5–23.

Alekseyev, E. Y. (1967). Do Yakuts have polyphony? [Есть ли у якутов многоголосие]. *Sovetskaya Muzyka* 5, 97–105. Available online at: http://eduard.alekseyev.org/work3.html

Alekseyev, E. Y. (1986). *Early Folkloric Intonation. Pitch Aspect [Раннефольклорное интонирование: звуковысотный аспект]*. Moscow: Soviet Composer. Available online at: http://eduard.alekseyev.org/rfi/index.html

Alekseyev, E. Y. (1988). *Folklore in the Context of Modern Culture: A Discourse About the Future of Folk Song [Фольклор в контексте современной культуры: рассуждения о судьбах народной песни]*. Moscow: Soviet Composer. Available online at: http://eduard.alekseyev.org/fic/index.html

Alekseyev, E. Y. (1991a). "Jaw Harp: undisclosed perspectives of research... [Варган: нераскрытые исследовательские перспективы]," in *Jew's Harp (khomus) and its music. Proceedings of the First All-Union Conference, 1988 [Варган (хомус) и его музыка. Материалы I Всесоюзной конференции 1988]*, ed E. Alekseyev (Yakutsk: Yakut Institute of language, literature and history of the USSR Academy of Science), 6–14. Available online at: http://eduard.alekseyev.org/work7.html

Alekseyev, E. Y. (1991b). "Jaw Harp that is playing, singing, talking... [Варган играющий, поющий, говорящий]," in *2nd International Jew's Harp Congress-Festival at Yakutsk* (Yakutsk: International Jew's Harp Society). Available online at: http://eduard.alekseyev.org/work35.html

Alekseyev, E. Y. (1993). *Speaking and Singing: Prolegomena to Anthropophonics [Пение и говорение. Основы антропофоники]*. Moscow: The Institute of History of Arts at the Ministry of Culture of the Russian Federation. Available online at: http://eduard.alekseyev.org/work42.html

Alekseyev, E. Y. (2008). "Singing in culturally-specific mental disorders of the Northern Siberian peoples [Пение при культурно-специфичных психических расстройствах у народов сибирского Севера]," in [Миф, символ, ритуал. Народы Сибири] *Myth, Symbol, Ritual. Peoples of Siberia*, ed S. Y. Nekliudov (Moscow: Russian State Humanitarian University), 95–107. Available online at: http://eduard.alekseyev.org/work4.html

Alekseyev, E. Y. (2016). The Yakut folk song: a brief ethnographic-musical sketch. *Anthropol. Archeol. Eurasia* 55, 41–59. doi: 10.1080/10611959.2016.1263484

Alekseyev, N. A. (1992). *Traditional Religious Beliefs of Turkic-Speaking Peoples of Siberia [Традиционные религиозные верования тюркоязычных народов Сибири]*. Novosibirsk: Nauka.

Alekseyev, G. G. (1986). "Khomus, play your song for them! [Заиграй им, хомус, свою песню]," in *Music of Russia: Musical Creativity and Musical Life in Republics of Russian Federation [Музыка России: музыкальное творчество и музыкальная жизнь республик Российской Федерации]*, ed A. Grigoryeva (Moscow: Soviet Composer), 321–332.

Alexeyev, I., and Shishigin, S. S. (2004). The music of the Yakut traditional khomus. *J. Int. Jews Harp Soc.* 1, 88–94.

Allen, E. J., and Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *J. Acous. Soc. Am.* 135, 1371–1379. doi: 10.1121/1.4863269

Anderson, D. G., and Bissett, T. G. (2015). "The initial colonization of north America: sea level change, shoreline movement, and great migrations," in *Mobility and Ancient Society in Asia and the Americas*, eds M. D. Frachetti and R. N. I. Spengler (Heidelberg: Springer International Publishing), 59–88. doi: 10.1007/978-3-319-15138-0_6

Andreyev, N. D., and Sunik, O. P. (1982). On the problem of relationship between Altaic languages and methods of its solution [О проблеме родства алтайских языков и методах ее решения]. *Voprosy Jazykoznanija* 2, 26–36.

Andreyeva, J. V. (1987). "Bronze age of far east [Бронзовый век Дальнего Востока]," in *Archaeology of the USSR. Bronze Age of the forest zone of the USSR [Археология СССР. Эпоха бронзы лесной полосы СССР]*, ed B. A. Rybakov (Moscow: Nauka), 351–357.

Antonov, N. K. (1997). "Yakut language [Якутский язык]," in *Languages of the World. Turkic Languages* [Языки мира: Тюркские языки], ed E. R. Tenishev (Bishkek: Russian Academy of Science, Kyrkyzstan), 17–34.

Antoshina, M. N. (1939). "Musical Development of children in children music school [Музыкальное развитие детей в детской музыкальной школе]," in *To a Music Teacher [Музыканту педагогу]*, ed G. Prokofiev (Moscow: Muzgiz), 25–44.

Anuchin, V. I. (1926). "The causes of population extinction in Northern Siberia [Причины вымирания населения в северной Сибири]," in *A Collection of Works Dedicated to Vladimir M. Bekhterev (for the 40th Anniversary of His Professorship [Сборник, посвященный Владимиру Михайловичу Бехтереву (к 40-летию профессорской деятельности)]*, ed V. M. Bekhterev (Leningrad: State Academy of Psychoneurology & State Institute of Brain Research), 649–661.

Arakelova, A. O. (2012). Professional Music Education during the Formative Period (1918-1930s) [Профессиональное музыкальное образование в период становления (1918-1930-е годы)]. *World Sci. Cult. Educ.* 2, 312–315.

Arcones, P. C. (2013). *Sons of Heaven, Brothers of Nature: The Naxi of Southwest China*. Kunming: Papers of the White Drago.

Arkhipov, N. D. (1989). *Ancient Cultures of Yakutia [Древние культуры Якутии]*. Yakutsk: Yakut Book Publishing.

ASA (1951). *American Standard Acoustical Terminology*, Vol. 25, eds L. Beranek, K. C. Morrical, and J. W. McNair. New York, NY: American Standards Association.

Atkinson, Q. D., Gray, R. D., and Drummond, A. J. (2009). Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc. R. Soc. B Biol. Sci.* 276, 367–373. doi: 10.1098/rspb.2008.0785

Austin, M. H. (1930). *The American Rhythm: Studies and Reexpressions of Amerindian Songs*. Boston, MA: Houghton Mifflin.

Avdeyev, A. D. (1960). "Mask. The experiment in typological classification according to ethnographic data [Маска. Опыт типологическои классификации по этнографическим материалам]," in *Collection of Works of the Museum of Anthropology and Ethnography [Сборник Музея антропологии и этнографии]*, ed L. P. Tolstov (Moscow; Leningrad: Museum of anthropology and ethnography), 39–110.

Baipakov, K. M., and Taimagambetov, Z. K. (2006). *Archaeology of Khazakhstan [Археология Казахстана]*. Almaty: Kazakh Universitety.

Bakx, P., and Crane, F. (2017). The 1000 Names of the Jew's harp. *Nomenclature of Over Thousand Proper Names for the Jew's Harp from All Over the World*, 28. Middelburg: Foundation Antropodium. Available online at: http://www.antropodium.nl/DuizendNamenMhpNOMENCLATUUR.htm (accessed January 1, 2019).

Balzano, G. J. (1986). What are musical pitch and timbre? *Music Percept. Interdiscipl. J.* 3, 297–314. doi: 10.2307/40285339

Barinova, E. (2013). *Ethnocultural Contacts of China with Ethnicities of Central Asia in Antiquity and Middle Ages [Этнокультурные контакты Китая с народами Центральной Азии в древности и средневековье]*. Moscow: Russian Academy of Science.

Barr, W. (1994). The Eighteenth Century Trade between the Ships of the Hudson's Bay Company and the Hudson Strait Inuit. *Arctic* 47, 236–246. doi: 10.14430/arctic1294

Barrett, M. (2011). Musical narratives: a study of a young child's identity work in and through music-making. *Psychol. Music* 39, 403–423. doi: 10.1177/0305735610373054

Bartlett, J. C., and Snelus, P. (1980). Lifespan memory for popular songs. *Am. J. Psychol.* 93, 551. doi: 10.2307/1422730

Bar-Yosef, O., Eren, M. I., Yuan, J., Cohen, D. J., and Li, Y. (2012). Were bamboo tools made in prehistoric Southeast Asia? An experimental view from South China. *Q. Int.* 269, 9–21. doi: 10.1016/j.quaint.2011.03.026

Beck, C. M., Deeds, E. E., Pozorski, S., and Pozorski, T. (1983). Pajatambo: an 18th century roadside structure in Peru. *Histor. Archaeol.* 17, 55–68. doi: 10.1007/BF03374031

Beliayev, V. M. (1933). *Musical Instruments of Uzbekistan [Музыкальные инструменты Узбекистана]*. Moscow: Gos Muz Izdat [Гос. музыкальное изд-во].

Beliayeva-Ekzempliarskaya, S. (1925). "Musical experience in preschool age [Музыкальное переживание в дошкольном возрасте]," in *Collection of Works of the Physiolo-psychological Department [Сборник работ физиолого-психологической секции]*, ed S. Beliayeva-Ekzempliarskaya (Moscow: The State Institute of Musical Science [Гос ин-та музыкальной науки.]), 3–29.

Bicheool, V. (2009). The current state of Nganasan culture [Современное состояние культуры нганасан]. *Herald Chelyabinsk State Acad. Cult. Arts* 4, 49–54.

Bjørkvold, J. R. (1992). *The Muse Within: Creativity and Communication, Song and Play from Childhood Through Maturity*. London: Harper Collins Publishers.

Blacking, J. (1967). *Venda Children's Songs: A Study in Ethnomusicological Analysis*. Chicago, IL: University of Chicago Press.

Blench, R. (2004). "Musical aspects of Austronesian culture," in *Proceedings of the Xth Association of Southeast Asian Archaeologists Conference, London 2004*, eds L. Bacus and I. Glover (Cambridge, UK: Kay Williamson Educational Foundation), 1–10.

Blium, D. (1977). "The role of dictation in development of professional musical hearing [Роль диктанта в развитии профессионального музыкального слуха]," in *Development of Musical Hearing [Воспитание музыкального слуха*, ed A. Agazhanov (Moscow: Muzyka [Музыка]), 86–117.

Bogoras, W. (1910). *On Psychology of Shamnism Amongst the Peoples of North-Eastern Asia [К психологии шаманства у народов Северо-Восточной Азии]*. Moscow: The Imperial Moscow University Publishing.

Bonnel, A.-M., Faita, F., Peretz, I., and Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. *Percept. Psychophys.* 63, 1201–1213. doi: 10.3758/BF03194534

Borodovskii, A. P. (2007). *Ancient Carved Horn From Southern Siberia of the Paleometallic Age [Древний резной рог Южной Сибири эпохи палеометалла]* (Dissertation). The Institute of Archaeology and Ethnography, The Novosibirsk State University, Novosibirsk, Russia.

Borodovskii, A. P. (2017). "Bone Jaw Harps and their workpieces of the Hunno-Sarmatian period in the North Altai region [Костяные варганы и их заготовки гунно-сарматского времени на территории Северного Алтая]," in *Problems of Archaeology, Ethnography, Anthropology of Siberia and Adjacent Territories [Проблемы археологии, этнографии, антропологии Сибири и сопредельных территорий]*, eds A. P. Derevianko and V. I. Molodin (Novosibirsk: The Institute of Archaeology and Ethnography of Russian Academy of Science), 279–283.

Brodianskii, D. L. (1978). "Yet another field of neolithic art in Far East [Еще одна область неолитического искусства на Дальнем Востоке]," in *At the Origin of Creativity. Primitive Art [У истоков творчества. Первобытное искусство]*, ed R. S. Vasilyevskii (Novosibirsk: Nauka), 133–141.

Brodsky, I. A. (1972). *Music of Peoples of Far East of the USSR [Музыка народов Дальнего Востока СССР] LP*. Moscow: Melodiya.

Bronson, B. (1996). "Metals, specialization, and development in Early Eastern and Southern Asia," in *Craft Specialization and Social Evolution: In Memory of V. Gordon Childe,* ed B. Wailes (Philadelphia, PA: The University Museum of Archaeology and Anthropology, University of Pennsylvania), 177–186.

Brown, S. (2007). Contagious heterophony: a new theory about the origins of music. *Musicae Scientiae* 11, 3–26. doi: 10.1177/1029864907011001 00101

Bulgakova, T. D. (2000). "Traditional views of the Nanai on music [Традиционные представления нанайцев о музыке]," in *Musical Ethnography of Tungus-Manchu Peoples. International Conference at Yakutsk [Музыкальная этнография тунгусо-маньчжурских народов]* (Yakutsk: The Institute of Problems of Exploration of North), 36–37.

Bulgakova, T. D. (2001). Music in traditional Nanai culture [Музыка в традиционной нанайской культуре]. *Inform. Bull. Russ. Found. Basic Res.* 3, 213–218.

Bunker, E. C. (1993). Gold in the ancient Chinese world: a cultural puzzle. *Artibus Asiae* 53, 27–50. doi: 10.2307/3250506

Buzhilova, A., Derevianko, A., and Shunkov, M. (2017). The northern dispersal route bioarchaeological data from the late Pleistocene of Altai, Siberia. *Curr. Anthropol.* 58, S491–S503. doi: 10.1086/694232

Caclin, A., Giard, M.-H., Smith, B. K., and McAdams, S. (2007). Interactive processing of timbre dimensions: a Garner interference study. *Brain Res.* 1138, 159–170. doi: 10.1016/j.brainres.2006.12.065

Campbell, P. S. (1998). *Songs in Their Heads: Music and Its Meaning in Children's Lives*. Oxford: Oxford University Press.

Canave-Dioquino, C., Santos, R. P., and Maceda, J. (2008). "The philippines," in *The Garland Handbook of Southeast Asian Music*, eds T. Miller and S. Williams (London: Routledge), 415–445.

Carlson, R., Granstrom, B., and Fant, G. (1970). Some studies concerning perception of isolated vowels. *STL-QPSR* 11, 19–35.

Chernykh, E. N. (1966). *History of the Earliest Metallurgy of Eastern Europe [История древнеишеи металлургии Восточнои Европы]*. Moscow: Nauka.

Chernykh, E. N. (1992). *Ancient Metallurgy in the USSR: The Early Metal Age*. Cambridge, UK: Cambridge University Press.

Chernykh, E. N. (2012). "The northern hunters, fishers and gatherers and their southern neighbours in Eurasia in the early metal age," in *Networks, Interaction and Emerging Identities in Fennoscandia and Beyond* (Helsinki: Société Finno-Ougrienne), 105–122. Available online at: https://www.sgr.fi/sust/sust265/sust265_chernykh.pdf (accessed February 14, 2019).

Christian, D. (2000). Silk roads or steppe roads? The silk roads in world history. *J. World History* 11, 1–26. doi: 10.1353/jwh.2000.0004

Chuluunbaatar, O. (2016). "Rare archaeological musical artefacts from ancient tombs in Mongolia," in *Studia Instrumentorum Musicae Popularis*, ed G. Jähnichen (Münster: MV Wissenschaft), 225–250.

Ciarla, R. (2007). Rethinking Yuanlongpo: the case for technological links between the Lingnan (PRC) and central Thailand in the Bronze Age. *East West* 57, 305–328. Available online at: https://www.jstor.org/stable/29757731

Coghlan, H. H. (1939). Prehistoric copper and some experiments in smelting. *Trans. Newcomen Soc.* 20, 49–65. doi: 10.1179/tns.1939.005.

Crabtree, D. E., and Davis, E. L. (1968). Experimental manufacture of wooden implements with tools of flaked stone. *Science* 159, 426–428. doi: 10.1126/science.159.3813.426.

Crane, F. (1982). Jew's (Jaw's? Jeu? Jeugd? Gew-gaw? Juice?) Harp. *Vierundzwanzigstel- jahrschrift der Internationalen Maultrommelvirtuosengenossenschaft* 1, 29–41.

Crane, F. (2007). Review of Jew's harps in European archaeology. *J. Int. Jews Harp Soc.* 4, 68–69.

Cross, I. (1999). "Is music the most important thing we ever did? Music, development and evolution," in *Music, Mind and Science*, ed S. W. Yi (Seoul: Seoul National University Press), 10–39.

Cross, I., and Morley, I. (2009). "The evolution of music: theories, definitions and the nature of the evidence," in *Communicative Musicality: Exploring the Basis of Human Companionship*, ed T. C. Malloch (London: Oxford University Press), 61–82.

Czaplicka, M. (1914). *Aboriginal Siberia, A study in Social Anthropology*. Oxford, UK: Clarendon Press.

Davis, B. L., and MacNeilage, P. F. (1995). The articulatory basis of babbling. *J. Speech Lang. Hear. Res.* 38, 1199–1211. doi: 10.1044/jshr.3806.1199

Davis, B. L., and Zajdo, K. (2008). *The Syllable in Speech Production: Perspectives on the Frame Content Theory*. Abingdon; Oxfordshire: Lawrence Erlbaum Associates.

de Ramón, I. A., and Rivera, I. A. (1982). Indigenous music of Venezuela. *World Mus.* 24, 22–37.

Devlet, M. A. (1976). *Petroglyphs of Ulug-Khem [Петроглифы Улуг-Хема]*. Moscow: Nauka.

Devlet, M. A. (1980). *Petroglyphs of Mugur-Sargol [Петроглифы Мугур-Саргола]*. Moscow: Nauka.

Devlet, Y. G., and Devlet, M. A. (2011). *The Treasures of Rock Art of North and Central Asia [Сокровища наскального искусства Северной и Центральной Азии]*. Moscow: Institute of Archaeology at Russian Academy of Science.

Dickinson, K. (2001). 'Believe'? Vocoders, digitalised female identity and camp. *Popular Mus.* 20, 333–347. doi: 10.1017/S0261143001001532

Dixon, E. J. (2015). "Late Pleistocene colonization of north America from northeast Asia: new insights from large-scale paleogeographic reconstructions," in *Mobility and Ancient Society in Asia and the Americas*, eds M. D. Frachetti and R. N. I. Spengler (Heidelberg: Springer International Publishing), 169–184. doi: 10.1007/978-3-319-15138-0_12

Dobzhanskaya, O. (1991). "Chukchi playing patterns on frame Jew's Harps [Чукотские наигрыши на рамном варгане]," in *Jew's Harp (khomus) and Its Music. Proceedings of the First All-Union Conference, 1988* [Варган (хомус) и его музыка. Материалы I Всесоюзной конференции 1988], ed E. Y. Alekseyev (Yakutsk: Yakut Institute of Language, Literature and History of the USSR Academy of Science), 52–58.

Dobzhanskaya, O. (2008). *Shamanic music of samoyed ethnoses in the syncretic unity of ritual* [*Шаманская музыка самодийских народов в синкретическом единстве обряда*] (Dissertation). The Moscow State Conservatory named after Tchaikovsky, The Department of Music Theory.

Dobzhanskaya, O. (2017). Personal songs of tundra Nenets of Taimyr [Личные песни тундровых ненцев Таймыра]. *Cult. Civil.* 7, 565–575.

Dorina, M. V. (2004). *Wind and percussive instruments of Sayano-Altaian Turks: An attempt of historic-ethnographic investigation [Духовые и Ударные Музыкальные Инструменты Тюрков Саяно-Алтая (Опыт Историко-Этнографического Исследования)]* (Dissertation). Tomsk University, Department of Ethnography and History of Science.

Dorokhova, Y. A. (1995). "Types of ritual vocal intonation and structure of traditional society [Виды ритуального вокального интонирования и структура традиционного сообщества]," in *Voice Abd Ritual. Proceedings of the Conference [Голос и ритуал. Материалы конференции. Май 1995]*, eds Y. A. Dorokhova, N. I. Zhulanova, and O. A. Pashina (Moscow), 89–93.

Dournon-Taurelle, G. (1975). "La Guimbarde," in *Paris Nanterre University* (Dissertation). Paris Nanterre University, Maison de l'archeologie et de l'ethnologie Rene Ginouves.

Dournon-Taurelle, G., and Wright, J. (1978). *Les Guimbardes du Musée de l'Homme*. Paris: Institut d'Ethnologie, Muséum national d'Histoire naturelle.

Dowling, W. J. (1984). Cognitive processes in the perception of art. *Adv. Psychol.* 19, 145–163. doi: 10.1016/S0166-4115(08)62350-X

Drabkin, W. (2001). "Theme," in *The New Grove Dictionary of Music and Musicians*, eds S. Sadie and J. Tyrrell (London: Macmillan Publishers). doi: 10.1093/gmo/9781561592630.article.27789

During, J., and During, E. (2015). "De l'espace lisse au temps troué: à propos des musiques nomades," in *Deleuze et la Musique*, eds P. Criton and J.-M. Chouvel (Paris: Centre de Documentation de la Musique Contemporaine (CDMC)), 327–330.

Duvan, N. D. (2000). "Personality of Ulch Shaman and his ritual [Личность ульчского шамана и его ритуал]," in *Musical Ethnography of Tungus-Manchu peoples. International Conference at Yakutsk [Музыкальная этнография тунгусо-маньчжурских народов]*, ed Y. Sheikin (Yakutsk: The Institute of Problems of Exploration of North), 51–53.

Duvan, N. D. (2003). "Musical instruments of Ulch people [Музыкальные инструменты ульчей]," in *Papers of Grodekovsky Museum*, eds N. Ruban and T. Melnikova (Khabarovsk: Khabarovsk regional folk museum named after Grodekovsky), 57–72.

Dyakonova, V. (2011). [*Изучение хомуса и его современное развитие в Якутии*] "Investigation of khomus and its contemporary development in Yakutia," in [Фольклор в контексте культуры] *Folklore in the Context of Culture*, ed M. Abdulayeva (Makhachkala: Dagestan State pedagogical University), 26–30.

Dyakonova, V. (2015). *Expedition to Chukotka Autonomous Region: Diary, Audio- and Video-Recordings, Photographs [Комплексная экспедиция в Чукотский автономный округ: дневники, аудио- и видеозаписи, фотографии]. Yakutsk.*

Dyakonova, V. (2017). *Musical instruments of Sakha people in light of classic typologies* [Музыкальные инструменты народа саха в свете классических типологий] (Dissertation). The Department of Art Studies, The Arctic State Institute of Art and Culture, Yakutsk, Russia.

Dyakonova, V. E., and Grigoryeva, V. G. (2017). "On the healing properties of sound in traditional Sakha culture [О целительных своиствах звука в традиционнои культуре саха]," in *Medical ethnography: contemporary approaches and conceptions [Медицинская этнография: современные подходы и концепции]*, ed V. I. Kharitonova (Moscow: Institute of Ethnography & Anthropology of Russian Academy of Science, OOO Publicity), 160–172.

Dyakonova, V. P. (1981). "Tuvan shamans and their social role in society [Тувинские шаманы и их социальная роль в обществе]," in *Problems of History of Social Consciousness of Siberian Aborigenes [Проблемы истории общественного сознания аборигенов Сибири]*, ed I. S. Vdovin (Leningrad: Nauka), 129–164.

Emsheimer, E. (1986). "Jew's harps in Siberia and Middle Asia [Варганы в Сибири и Средней Азии]," in *Problems of Traditional Instrumental Music of Peoples of USSR: Instrument-Performer-Music [Проблемы традиционной инструментальной музыки народов СССР: Инструмент-исполнитель-музыка]*, ed I. Matsiyevskii (Leningrad: LGITMiK), 80–94.

Endovitskaya, T. V. (1959). "On the pitch discriminatory sensitivity in children of the preschool age [О звуковысотной различительной чувствительности у детей дошкольного возраста]," in *Reports of the Academy of the Academy of Pedagogical Sciences* (Leningrad: Academy of Pedagogical Sciences of Russia), 42–46.

Endovitskaya, T. V. (1964). "Development of sensation and perception in the preschool age children [Развитие ощущения и восприятия у детей дошкольного возраста]," in *Psychol. Presch. Age Child. [Психология детей*

*дошкольного возраста]*, eds A. V Zaporozhets and D. B. Elkonin (Moscow: Prosvesheniye), 13–71.

Erb-Satullo, N. L. (2011). "Alloys from anau: the manipulation of metallic properties in 3rd Millennium B.C. Southern Central Asia," *in Materials Issues in Art and Archaeology MRS Symposium Proceedings,* eds P. Vandiver, W. Li, S. Ruvalcaba, C. L. Reedy, and L. D. Frame (Cambridge: Cambridge University Press), 157–163. doi: 10.1557/opl.2011.739

Everstova, A. V., Afanasyeva, D. N., and Arzhakova, S. N. (2019). "Jaw harp therapy and throat singing as a method of rehabilitation of children with delay in speech development [Хомусотерапия и горловое пение как метод восстановительного лечения детей с задержкой речевого развития]," in *Health as a Resource - v.2 [Здоровье как ресурс - 2]* ed C. A. Markov (Nizhnii Novgorod: Nizhegorodskii State University named after Lobachevsky).

Fales, C. (2002). The paradox of timbre. *Ethnomusicology* 46, 56–95. doi: 10.2307/852808

Feodorov, G. B. (1954). The results of 3-year work in Moldovia in Slavic-Russian archaeology [Итоги трёхлетних работ в Молдавии в области славяно-русской археологии]. *Brief Commun. Instit. Archaeol.* 56, 8–23.

Fernald, A. (1992). "Meaningful melodies in mothers' speech to infants," in *Nonverbal Vocal Communication Comparative and Developmental Approaches*, eds H. Papousek, U. Jurgens, and M. Papousek (Cambridge, UK: Cambridge University Press), 262–282.

Fischer, H. (1986). *Sound-Producing Instruments in Oceania: Construction and Playing Technique - Distribution and Function*, ed D. Niles. Boroko: Institute of Papua New Guinea Studies.

Fischer-Jørgensen, E. (1978). On the universal character of phonetic symbolism with special reference to vowels. *Studia Linguistica* 32, 80–90. doi: 10.1111/j.1467-9582.1978.tb00329.x

Fomin, V. P. (2018). Children phono-instruments of Southern Altaians [Детские фоноинструменты южных алтаицев]. *World Sci. Cult. Educ.* 2, 373–375.

Foster, M. L. (1994). "Symbolism: the foundation of culture," in *Companion Encyclopedia of Anthropology*, ed T. Ingold (New York, NY: Routledge), 366–395.

Foulks, E. F. (1985). "The transformation of arctic hysteria," in *The Culture-Bound Syndromes* (Dordrecht: Springer), 307–324. doi: 10.1007/978-94-009-5251-5_28

Fox, L. (1988). *The Jew's Harp: A Comprehensive Anthology, 2nd Edn*. London, Lewisburg PA: Associated University Presses, Bucknell University Press.

Fox, L. (1994). Kulturgeschichte der Maultrommel. *Notes* 50, 927–929. doi: 10.2307/898542

Francfort, H.-P. (1994). The Central Asian dimension of the symbolic system in Bactria and Margiana. *Antiquity* 68, 406–418. doi: 10.1017/S0003598X00046755

Frayonov, V. P. (1981). "Texture [фактура]," in *Encyclopedia of Music [Музыкальная энциклопедия]*, 754–761.

French, J. C. (2016). Demography and the palaeolithic archaeological record. *J. Archaeol. Method Theory* 23, 150–199. doi: 10.1007/s10816-014-9237-4

Funk, D. A. (2006). "Economic activities and materialized culture [Хозяйственные занятия и материализованная культура]," in *Turkic Peoples of Siberia [Тюркские народы Сибири]*, eds D. A. Funk and N. A. Tomilov (Moscow: Nauka), 186–204.

Gadzhiyeva, N. E. (1997). "Turkic languages [Тюркские языки]," in *Languages of the World. Turkic Languages [Языки мира: Тюркские языки]*, ed E. R. Tenishev (Bishkek: Russian Academy of Science), 17–34.

Galaiskaya, R. B. (1973). "Jaw Harp by the ethnicities of the Soviet Union: on the issue of archaism and folk music instruments [Варган у народов Советского Союза: к вопросу об архаизмах и народном музыкальном инструментарии]," in *Problems of Musical Folklore of the USSR Peoples: Articles and Data*, ed I. I. Zemtsovskii (Moscow: Muzyka [Музыка]), 328–350.

Galaiskaya, R. B. (1987). The Jew's Harp among the Peoples of the Soviet Union. *Vierundzwangzigsteljahresschrift der Internationalen Maultrommelvirtuosengenossensch* 3, 11–38.

Galdanova, G. R. (1987). *Pre-lamaistic Beliefs of Byriats [Доламаистские верования бурят]*. Novosibirsk: Nauka.

Gamov, I. I. (1894). *Sketches on Distant Siberia [Очерки далекой Сибири]*. Gomel: Ya. G. Syrkin Publishing.

Garfias, R. (1990). "An ethnomusicologist's thoughts on the processes of language and music acquisition," in *Music and Child Development: Proceedings of the 1987 Denver Conference*, eds F. Wilson and F. Roehmann (Ann Arbor, MI: Book Crafters), 100–105. Available online at: https://dialnet.unirioja.es/servlet/articulo?codigo=558324 (accessed July 17, 2018).

Garner, J. (2015). "Bronze Age tin mines in central Asia, in *Archaeometallurgy in Europe III. Proceedings of the 3rd International Conference Deutsches Bergbau-Museum Bochum*, eds A. Hauptmann and D. Modarressi-Tehrani (Bochum: Springer), 135–144.

Gelegdorj, E. (2015). "Copper exploitation, ore dressing and bronze melting activities of ancient habitants of Mongolia [К вопросу о добыче, обогащении и плавке меди древними жителями Монголии]," in *Ancient Metallurgy of the Sayan-Altai and East Asia*, eds Y. Murakami and Y. Esin (Abakan: Ekhime University Press), 25–34.

Gemuyev, I. N. (2000). *Mansi People: The Embodiment of a Myth [Народ манси: воплощение мифа]*. Novosibirsk: Institute of archaeology and ethnography.

Gemuyev, I. N., and Sagalayev, A. M. (1986). *Religion of Mansi People. Cult Locations of the 19th-early 20th Centuries [Религия народа манси. Культовые места XIX — начала XX в]*. Novosibirsk: Nauka.

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *J. Acous. Soc. Am.* 54, 1496–1516. doi: 10.1121/1.1914448

Goltsova, T. V., Osipova, L. P., Zhadanov, S. I., and Villems, R. (2005). The effect of marriage migration on the genetic structure of the Taimyr Nganasan population: genealogical analysis inferred from MtDNA markers. *Russ. J. Genet.* 41, 779–788. doi: 10.1007/s11177-005-0160-3

Golubkova, A. N., and Ivanov, A. G. (1997). Ancient jaw harps of kama area [Древние Варганы Прикамья]. *Courier Udmurt Univ.* 8, 93–107.

Grachiova, G. N. (1983). *Traditional Worldview of the Taimyr Hunters (based on Nganasan data, XIX-XX centuries) [Традиционное мировоззрение охотников Таймыра (на материалах нганасан XIX–XX века)]*. Leningrad: Nauka.

Green, M., and Hodge, P. (1980). *Inuit Throat And Harp Songs, Eskimo Women's Music of Povungnituk*. Toronto, ON: The Canadian Music Heritage Collection.

Grigoryan, G. A. (1957). "Music culture of Yakut ASSR [Музыкальная культура Якутской АССР]," in *Music Culture of the Autonomous Republics of the Russian Federation [Музыкальная культура автономных республик РСФСР]*, ed G. I. Litinskii (Moscow: Muzgiz), 331–348.

Grigoryeva, A. M. (1996). *Folk Healing of the People in Yakutia of the XVII-XX Centuries [Народное врачевание в Якутии: XVII-XX вв.]*. Moscow: Institute of archaeology and ethnography of Russian Academy of Science.

Grimault, N., Micheyl, C., Carlyon, R. P., and Collet, L. (2002). Evidence for two pitch encoding mechanisms using a selective auditory training paradigm. *Percept. Psychophys.* 64, 189–197. doi: 10.3758/BF03195785

Grove, M. (2012). "The archaeology of group size," in *Social Brain, Distributed Mind*, eds R. Dunbar, C. Gamble, and J. Gowlett (London: British Academy), 391–411.

Grove, M., Pearce, E., and Dunbar, R. I. M. (2012). Fission-fusion and the evolution of hominin social systems. *J. Hum. Evol.* 62, 191–200. doi: 10.1016/j.jhevol.2011.10.012

Haid, G. (1999). The trump and eroticism. *Vie rundzwanzigsteljahrsschrift der Internationalen Maultrommelvirtuosengenossenschaft* 8, 60–80.

Hais, M., Komprdová, K., Ermakov, N., and Chytrý, M. (2015). Modelling the Last Glacial Maximum environments for a refugium of Pleistocene biota in the Russian Altai Mountains, Siberia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 438, 135–145. doi: 10.1016/j.palaeo.2015.07.037

Halpern, A. R., and Bartlett, J. C. (2010). "Memory for Melodies," in *Music Perception*, eds M. R. Jones, R. R. Fay, and A. N. Popper (New York, NY: Springer), 233–258. doi: 10.1007/978-1-4419-6114-3_8

Halpern, A. R., and Bartlett, J. C. (2011). The persistence of musical memories: a descriptive study of earworms. *Music Percept. Interdiscipl. J.* 28, 425–432. doi: 10.1525/mp.2011.28.4.425

Hartley, J. M. (2014). *Siberia: A History of the People*. Cambridge, MA: Yale University Press. doi: 10.12987/yale/9780300167948.001.0001

Hassan, F. A. (1981). *Demographic Archaeology*. New York, NY: Academic Press. doi: 10.1016/B978-0-12-624180-8.50010-X

Hauptmann, A. (2007). *The Archaeometallurgy of Copper. Evidence from Faynan, Jordan.* Berlin: Springer. doi: 10.1007/978-3-540-72238-0

Hauser-Schäublin, B. (1995). "Puberty rites, women's Naven, and initiation: women's rituals of transition in Abelam and Iatmul culture," in *Gender Rituals: Female Initiation in Melanesia*, eds N. C. Lutkehaus and P. B. Roscoe (London: Routledge), 33–53.

Herva, V.-P., Ikäheimo, J., Kuusela, J.-M., and Nordqvist, K. (2009). "Close encounters of the copper kind," in *Proceedings of the 10th Nordic TAG Conference: Stiklestad, Norway*, eds R. Berge, M. E. Jasinski, and K. Sognnes (London: Britsh Archaeological Reports), 109–115.

Higham, C., Higham, T., Ciarla, R., Douka, K., Kijngam, A., and Rispoli, F. (2011). The origins of the bronze age of Southeast Asia. *J. World Prehistory* 24, 227–274. doi: 10.1007/s10963-011-9054-6

Hoffecker, J. F., and Elias, S. A. (2007). *Human Ecology of Beringia*. New York, NY: Columbia University Press. doi: 10.7312/hoff13060

Honeychurch, W. (2015). *Inner Asia and the Spatial Politics of Empire: Archaeology, Mobility, and Culture Contact*. New York, NY: Springer. doi: 10.1007/978-1-4939-1815-7

Hope, G., Kershaw, A. P., van der Kaars, S., Xiangjun, S., Liew, P. M., Heusser, L. E., et al. (2004). History of vegetation and habitat change in the Austral-Asian region. *Q. Int.* 118–119, 103–126. doi: 10.1016/S1040-6182(03)0 0133-2

Houtsma, A. J. M., and Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *J. Acous. Soc. Am.* 87, 304–310. doi: 10.1121/1.399297

Hsu, T.-H. (2001). "Taiwan: music of the Taiwan aborigines," in *The Garland Encyclopedia of World Music, Volume 7, East Asia: China, Japan and Korea*, eds R. C. Provine, Y. Tokumaru, and L. Witzleben (London: Routledge), 523–529.

Hung, H., and Chao, C. (2016). Taiwan's early metal age and southeast Asian trading systems. *Antiquity* 90, 1537–1551. doi: 10.15184/aqy.2016.184

Huo, W. (2016). On the early metal objects and early Metal Age in Tibet. *Chin. Archaeol.* 16, 91–96. doi: 10.1515/char-2016-0008

Huron, D. (2001). Tone and voice: a derivation of the rules of voice-leading from perceptual principles. *Music Percept.* 19, 1–64. doi: 10.1525/mp.2001.19.1.1

Ishi, S. (1916). "The Island of Formosa and its primitive inhabitants," in *Transactions and Proceedings of the Japan Society London,* ed H. L. Joly (London: Kegan Paul, Trench, Trübner and Co), 38–60.

Ivanov, S. V. (1975). *Masks of Siberian Peoples [Маски народов Сибири]*. Leningrad: Avrora.

Jaago, C. (2009). *Estonian Jew's Harp Stories: The Reproduction and Restoration of Melody-Play Tradition.* Available online at: https://brage.bibsys.no/xmlui/handle/11250/2439116 (accessed June 13, 2018).

Jaffe-Berg, E. (2001). Forays into grammelot the language of nonsense. *J. Dramat. Theor. Crit.* 2, 3–16.

Jochelson, W. (1910). *The Yukaghir and the Yukaghirized Tungus*. Leiden: E. J. Brill ltd.; G.E. Stechert.

Jochelson, W. (1928). *Peoples of Asiatic Russia.* Washington DC: The American Museum of Natural History. doi: 10.5962/bhl.title.68269

Jordania, J. (2006). *Who Asked the First Question? The Origins of Human Choral Singing, Intelligence, Language and Speech*. Tbilisi, GA: Logos.

Kalinichenko, E. P., and Alekseyev, M. A. (2009). "Meditative-imagery musical (Jaw Harp) psychotherapy as a method of medical rehabilitation of patients with posttraumatic stress disorders. [Медитативно-образная музыкальная (варганная) психотерапия как метод медицинской реабилитации пациентов с посттравматическими стрессовыми реакциями," in *Achievements and Perspectives of Medical Rehabilitation: Proceedings of the Conference Dedicated to 75 Year Anniversary of the Air Force Sanatorium at Chemitokvadzhe* [Достижения и перспективы медицинской реабилитации: материалы научно-практической конференции, 4-6 июня 2009 года, г. Сочи, военный санаторий Чемитоквадже], ed V. V. Yamenskov (Sochi: Federal State Department), 338–340.

Kaniuth, K. (2007). The Metallurgy of the Late Bronze Age Sapalli Culture (Southern Uzbekistan) and its implications for the "tin question." *Iranica Antiqua* 42, 23–40. doi: 10.2143/IA.42.0.201 7869

Kartomi, M. J. (2012). *Musical Journeys in Sumatra*. Champaign, IL: University of Illinois Press. doi: 10.5406/illinois/9780252036712.001.0 001

Kartomi, M. J., Sutton, R. A., Suanda, E., Williams, S., and Harnish, D. (2008). "Indonesia," in *The Garland Handbook of Southeast Asian Music*, eds T. Miller and S. Williams (London: Routledge), 334–405.

Kendall, R. A., and Carterette, E. C. (1993). Identification and blend of timbres as a basis for orchestration. *Contemp. Music Rev.* 9, 51–67. doi: 10.1080/07494469300640341

Kholopova, V. (1979). *Texture: An Essay [Фактура: Очерк]*. Moscow: Muzyka [Музыка].

Kiriushin, Y. F. (2002). *Eneolithic and early Bronze Ages of south of Eastern Siberia [Энеолит и ранняя бронза юга Западной Сибири]*. Barnaul: The Altai University Publishing.

Kirmayer, L. J. (2018). "Ethno- and cultural psychiatry," in *The International Encyclopedia of Anthropology*, Vol. 11, ed H. Callan (Hoboken, NJ: Wiley). doi: 10.1002/9781118924396.wbiea2362

Kiyekbayev, J. G. (1972). *Introduction to Uralo-Altaic Linguistics [Введение в урало-алтаиское языкознание]*. Ufa: Bashkir Book Publishing.

Ko, S., Joseph, A., and Whitman, J. B. (2014). "Comparative consequences of the tongue root harmony analysis for proto-Tungusic, proto-Mongolic, and proto-Korean," in *Paradigm Change: In the Transeurasian Languages and Beyond*, eds M. Robbeets and W. Bisang (Amsterdam: John Benjamins), 141–176. doi: 10.1075/slcs.161.13ko

Kochmar, N. N. (1994). *Petroglyphs of Yakutia. History and Culture of the East of Asia [Писаницы Якутии. История и культура Востока Азии]*. Novosibirsk: The Institute of Archaeology and Ethnography of Russian Academy of Science.

Koizumi, F., Tokumaru, Y., and Yamaguchi, O. (1977). *Asian Musics in an Asian Perspective.* Report of Asian Traditional Performing Arts. Tokyo: Heibonsha Limited Publishers.

Kolltveit, G. (2006). *Jew's Harps in European Archaeology*. Oxford UK: Archaeopress.

Kolltveit, G. (2009). The Jew's Harp in Western Europe: trade, communication, and innovation, 1150–1500. *Yearbook Trad. Music* 41, 42–61. Available online at: https://www.jstor.org/stable/25735478

Kolltveit, G. (2016). "Jew's harps of bone, wood and metal how to understand construction, classification and chronology," in *Studien zur Musikarchäologie*, eds R. Eichmann, J. Fang, and L.-C. Koch (Rahden: Verlag Marie Leidorf GmbH), 63–73.

Kon'kova, L. V. (1989). *Bronze Smelting Production in the South of the USSR Far East, the Turn of II-I m.BC to XIII Century AD [Бронзолитейное производство на юге Дальнего Востока СССР, рубеж II-I тыс. до н. э.-XIII век н. э]*. Leningrad: Nauka.

Koops, L. H. (2012). "Now can I watch my video?": exploring musical play through video sharing and social networking in an early childhood music class. *Res. Stud. Mus. Educ.* 34, 15–28. doi: 10.1177/1321103X12442994

Koskoff, E. (2008). *The Concise Garland Encyclopedia of World Music: The Middle East, South Asia, East Asia, Southeast Asia*. London: Routledge.

Kreinovich, E. A. (1928). *A Diary [Дневник]*. Yuzhno-Sakhalinsk: Sakhalin Regional Museum of Natural Science.

Krumhansl, C. L. (1989). "Why is musical timbre so hard to understand," in *Perception of Electroacoustic Sound and Music*, eds S. Nielzén and O. Olsson (New York, NY: Excerpta Medica), 43–54.

Krushanov, A. I. (1987). *History and Culture of Chukchi: Historic-Ethnographic Sketches [История и культура чукчей: историко-этнографические очерки]*. Leningrad: Nauka.

Kunanbayeva, A. (1987). A living legend. *Vie rundzwanzigsteljahrsschrift der Internationalen Maultrommelvirtuosengenossenschaft* 3, 66–69.

Kungurov, S. N. (1994). *Udmurt Traditional Musical Instruments [Удмуртские традиционные музыкальные инструменты]*. Izhevsk: Ministry of Culture of Udmurt Republic.

Kuzmina, E. E. (2008). *The Prehistory of the Silk Road.* Philadelphia, PA: University of Pennsylvania Press. doi: 10.9783/9780812292336

Kypchakova, L. V. (2006). On the matter of the Tree Cult of Altaians [К вопросу о культе деревьев у алтаицев]. *Siberian Pedag. J.* 3, 130–136.

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge, UK: Cambridge University Press.

Lahr, M. M., and Foley, R. A. (2001). "Genes, fossils and behavior: when and where do they fit?" in *Genes, Fossils, and Behaviour: An Integrated Approach to Human Evolution*, eds P. Donnelly and R. Foley (Amsterdam: IOS Press Ohmsha, NATO Science), 13–48.

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: evidence from eye-tracking. *J. Exp. Child Psychol.* 161, 32–45. doi: 10.1016/j.jecp.2017.03.017

Layton, R., and O'Hara, S. (2012). "Human social evolution: a comparison of hunter-gatherer and chimpanzee social organization," in *Social Brain, Distributed Mind*, eds R. Dunbar, C. Gamble, and J. Gowlett (London: British Academy), 83–113.

Le Roux, C. C. F. M. (1950). *De Bergpapoea's van Nieuw - Guinea en hun Woongevied*. Koninklijk. Leiden: Brill.

Lebedeva, G. C., and Kuhl, P. K. (2010). Sing that tune: infants' perception of melody and lyrics and the facilitation of phonetic recognition in songs. *Infant Behav. Dev.* 33, 419–430. doi: 10.1016/j.infbeh.2010.04.006

Ledang, O. K. (1972). On the acoustics and the systematic classification of the jaw's harp. *Yearbook Int. Folk Music Council* 4, 95–103.

Lee, G. Y., and Kisilevsky, B. S. (2014). Fetuses respond to father's voice but prefer mother's voice after birth. *Dev. Psychobiol.* 56, 1–11. doi: 10.1002/dev.21084

Leipp, É. (1963). Un ≪vocoder≫ mécanique: La guimbarde. *Annales des Télécommunications* 18, 82–87.

Leipp, É. (1967). La Guimbarde. *Bulletin du Groupe d'Acoustique Musical Faculte des Sciences Paris* 25, 25.

Lembke, S. A., and McAdams, S. (2015). The role of spectral-envelope characteristics in perceptual blending of wind-instrument sounds. *Acta Acust. United Acust.* 101, 1039–1051. doi: 10.3813/AAA.918898

Leontyev, A. N. (2009). *The Development of Mind. Selected Works of Aleksei Nikolaevich Leontyev,* ed M. Cole (Transl. M. Kipylova). Kettering, OH: Bookmasters, Inc., 115–136.

Leontyev, N. V. (1978). "Anthropomorphic images of Okunevskaya culture [Антропоморфные изображения окуневской культуры]," in *Siberia, Central and Eastern Asia in Antiquity: Neolithic and Age of Metal [Сибирь, Центральная и Восточная Азия в древности. Неолит и эпоха металла]* (Novosibirsk: Nauka), 88–118.

Leontyev, N. V., Kapelko, V. F., and Yesin, Y. N. (2006). *Statues and Stelae of the Okunev Culture [Изваяния и стелы окуневскои культуры]*. Abakan: Khakas Publishing House.

Leshchenko, N. V., and Prokopets, S. P. (2015). Medieval musical instruments (based on data of Primorye artifacts [Средневековые музыкальные инструменты (по материалам памятников Приморья)]. *Russ. Pac.* 4, 222–235.

Levin, T. C., and Suzukei, V. (2006). *Where Rivers and Mountains Sing: Sound, Music, and Nomadism in Tuva and Beyond*. Bloomington, IN: Indiana University Press.

Levinson, S. C. (2013). "Cross-cultural universals and communication structures," in *Language, Music, and the Brain: A Mysterious Relationship*, ed M. A. Arbib (Cambridge, MA: MIT Press), 67–82. doi: 10.7551/mitpress/9780262018104.003.0003

Lewis, J. (2013). "A cross-cultural perspective on the significance of music and dance to culture and society: insight from BaYaka pygmies," in *Language, music, and the brain: a mysterious relationship*, ed M. A. Arbib (Cambridge, MA: MIT Press), 45–66. doi: 10.7551/mitpress/9780262018104.003.0002

Li, H. (1956). A comparative study of the Jew's Harps among the Aborigines of Formosa and East Asia. *Bull. Institute of Ethnology, Academia Sinica* 1, 85–140.

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia* 7, 128–134. doi: 10.1007/BF02156143

Lisina, M. I. (1966). "Development of the cognitive capacity in children during their first half a year of life [Развитие познавательной деятельности детей первого полугодия жизни]," in *Development of Perception in Early and Preschool Childhood. [Развитие восприятия в раннем и дошкольном детстве]*, eds A. V Zaporozhets and M. I. Lisina (Moscow: Prosvesheniye), 16–48.

Loewy, J. V. (1995). The musical stages of speech: a developmental model of pre-verbal sound making. *Music Ther.* 13, 47–73. doi: 10.1093/mt/13.1.47

Lotman, Y. M. (2001). *Universe of the Mind: A Semiotic Theory of Culture*. London; New York, NY: I.B. Tauris.

L'vova, E. E., Oktiabr'skaya, I. V., and Sagalayev, A. M. (1989). *The Traditional Worldview of Turks of Southern Siberia. Humanbeing. Society [Традиционное мировоззрение тюрков Южнои Сибири. Человек. Общество]*. Novosibirsk: Nauka.

Maddox, A. (2009). "Rhetoric as a guide to vocal timbre and sonority in Italian recitative," in *Musical Islands: Exploring Connections between Music, Place and Research,* eds E. Mackinlay, B.-L. Bartleet, and K. Barney (Newcastle: Cambridge Scholars Publishing), 381–403.

Maier, A. (2017). Population and Settlement Dynamics from the Gravettian to the Magdalenian. *Mitteilungen der Gesellschaft fuer Urgeschichte* 26, 83–102.

Malinsky-Buller, A., and Hovers, E. (2019). One size does not fit all: group size and the late middle Pleistocene prehistoric archive. *J. Hum. Evol.* 127, 118–132. doi: 10.1016/j.jhevol.2018.11.002

Malloch, S. (2000). Mothers and infants and communicative musicality. *Mus. Sci.* 3, 29–57. doi: 10.1177/102986490000 30S104

Malloch, S., and Trevarthen, C. (2009). "Musicality: communicating the vitality and interests of life," in *Communicative Musicality: Exploring the Basis of Human Companionship*, eds S. Malloch and C. Trevarthen (New York, NY: Oxford University Press), 1–11.

Mamcheva, N. A. (2005). Nivkh Jew's Harps [Нивхские варганы]. *Courier Sakhalin Museum* 12, 271–284.

Mamcheva, N. A. (2012). *Musical Instruments in Nivkh Traditional Culture [Музыкальные инструменты в традиционной культуре нивхов]*. Yuzhno-Sakhalinsk: GUP Sakhalinskaya Regional Press.

Martinón-Torres, M., Wu, X., de Castro, J. M. B., Xing, S., and Liu, W. (2017). Homo sapiens in the Eastern Asian Late Pleistocene. *Curr. Anthropol.* 58, 434–448. doi: 10.1086/694449

Maskarinec, G. G. (1995). *The Rulings of the Night: An Ethnography of Nepalese Shaman Oral Texts*. Madison, WI: University of Wisconsin Press.

Maslov, A. L. (1911). *Illustrated Description of Musical Instruments from Dashkovskii Ethnographic Museum in Moscow [Иллюстрированное описание музыкальных инструментов, хранящихся в Дашковском Этнографическом Музее в Москве]*. Moscow: The society of connoisseurs of Natural Sciences, Anthropology and Ethnography.

Maturana, H. R., and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Dordrecht: Springer. doi: 10.1007/978-94-009-8947-4

Matusky, P. (2008). "Borneo: Sabah, Sarawak, Brunei, and Kalimantan," in *The Garland Handbook of Southeast Asian Music*, eds T. Miller and S. Williams (London: Routledge), 406–414.

Maykapar, S. (1900). *Musical Hearing, its Significance, Specialty and Method of Correct Development [Музыкальный слух. Его значение, природа, особенности и метод правильного развития]*. Moscow: Jurgenson [Юргенсон].

Mazel, L. (1979). *Structuring of Musical Works [Строение музыкальных произведений], 2nd Edn.* Moscow: Muzyka [Музыка].

Mazepus, V. V., and Galitskaya, S. P. (1997). *Musical Culture of Siberia. Traditional Culture of Indigenous People of Siberia* [*Музыкальная Культура Сибири. Традиционная Культура Коренных Народов Сибири*], Vol. 1. ed B. A. Shindin. Novosibirsk: Novosibirsk State Conservatory named after Glinka.

Mazhitov, N. A. (1981). *Burial Mounds of Southern Urals VIII-XII Centuries [Курганы Южного Урала VIII-XII вв.]*. Moscow: Nauka.

McAdams, S. (2019). "Timbre as a structuring force in music," in *Timbre: Acoustics, Perception, and Cognition,* eds K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, and R. R. Fay (Cham: Springer), 211–244. doi: 10.1007/978-3-030-14832-4_8

McAdams, S., and Bertoncini, J. (1997). Organization and discrimination of repeating sound sequences by newborn infants. *J. Acous. Soc. Am.* 102, 2945–2953. doi: 10.1121/1.420349

McGovern, J. B. M. (1922). *Among the Head-Hunters of Formosa*. London: T. Fischer Unwin Ltd.

McLean, M. (1999). *Weavers of Song: Polynesian Music and Dance*. Honolulu, HI: University of Hawai'i Press.

McPhee, C. (1955). "Children and Music in Bali," in *Childhood in Contemporary Cultures*, eds M. Mead and M. Wolfenstein (Chicago, IL: University of Chicago Press), 70–98.

Mei, J., Wang, P., Chen, K., Wang, L., Wang, Y., and Liu, Y. (2015). Archaeometallurgical studies in China: some recent developments and challenging issues. *J. Archaeol. Sci.* 56, 221–232. doi: 10.1016/j.jas.2015.02.026

Melara, R. D., and Marks, L. E. (1990a). Interaction among auditory dimensions: timbre, pitch, and loudness. *Percept. Psychophys.* 48, 169–178. doi: 10.3758/BF03207084

Melara, R. D., and Marks, L. E. (1990b). Perceptual primacy of dimensions: support for a model of dimensional interaction. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 398–414. doi: 10.1037/0096-1523.16.2.398

Menn, L., and Vihman, M. M. (2011). "Features in child phonology," in *Where Do Phonological Features Come from? Cognitive, Physical and Developmental Bases of Distinctive Speech Categories*, eds G. Nick Clements and R. Ridouane (Amsterdam; Philadelphia, PA: John Benjamins Publishing), 261–301. doi: 10.1075/ijcl.18.1.07odo

Merriam, A. P. (1964). *The Anthropology of Music*. Evanston, IL: Northwestern University Press.

Mersenne, M. (1957). *Harmonie Universelle: the Books on Instruments*. Leiden: Martin Nijhoff.

Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222–226. doi: 10.1126/science.1224344

Miller, T., and Williams, S. (2008). "Southeast asian musics: an overview," in *The Garland Handbook of Southeast Asian Music*, eds T. Miller and S. Williams (London: Routledge), 4–20.

Minks, A. (2002). From children's song to expressive practices: old and new directions in the ethnomusicological study of children. *Ethnomusicology* 46, 379–408. doi: 10.2307/852716

Mishra, T. N. (1994). The archaeological research in the high mountains of mustang district. *Ancient Nepal* 136, 147–160.

Mitskevich, S. I. (1929). *Menerik and Emiriacheniye. Forms of hysteria in Kolyma region [Мэнерик и эмирячение. Формы истерии в Колымском крае]*. Leningrad: Academy of Sciences of USSR.

Modorov, I. S., and Dvornikov, E. P. (2015). A historic essay on culture of ethnicities of Mountain Altai during the late 17th-20th centuries [Исторический очерк по культуре народов Горного Алтая в конце XVII - XX в.]. *World Eurasia* 4, 24–50.

Moog, H. (1976). *The Musical Experience of the Pre-school Child*. London: Schott Music.

Moore, B. C. J. (2012). *An Introduction to the Psychology of Hearing, 6th Edn*. Bingley: Emerald Group Publishing.

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *J. Acous. Soc. Am.* 77, 1853–1860. doi: 10.1121/1.391936

Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acous. Soc. Am.* 80, 479–483. doi: 10.1121/1.394043

Moore, B. C. J., Glasberg, B. R., and Shailer, M. J. (1984). Frequency and intensity difference limens for harmonics within complex tones. *J. Acous. Soc. Am.* 75, 550–561. doi: 10.1121/1.390527

Morgan, D. A. (2008). *Organs and Bodies: The Jew's Harp and the Anthropology of Musical Instruments* (Dissertation). The Department of Music, The University of British Columbia, Vancouver, Canada. doi: 10.14288/1.0066561

Morgan, D. A. (2017). *Speaking in Tongues: Music, Identity, and Representation in Jew's Harp Communities.* Available at: http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.729603 (accessed June 13, 2018).

Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology, and the Origins of Musicality*. Oxford: Oxford University Press. doi: 10.1093/acprof:osobl/9780199234080.001.0001

Mukhina, T. K., and Lisina, M. I. (1966). "The dependency of age and individual achievements in discriimination of pitch from the type of activity iin preschool age children [Зависимость возрастных и индивидуальных показателей звуковысотного дифференцирования от характера деятельности детей в пред," in *Development of Perception in Early and Preschool Childhood [Развитие восприятия в раннем и дошкольном детстве]*, eds A. V. Zaporozhets and M. I. Lisina (Moscow: Prosvesheniye), 49–73.

Nadeliayev, V. M. (1960). *The Project of Universal Unified Phonetic Transcription (UUPT) [Проект универсальной унифицированной фонетической транскрипции (УУФТ)]*. Moscow: Nauka.

Nadeliayev, V. M. (1980). "Articulatory classification of vowels [Артикуляционная классификация гласных]," in *Phonetic studies of Siberian languages [Фонетические исследования по сибирским языкам]*, ed V. M. Nadeliayev (Novosibirsk: Nauka), 3–91.

Nattiez, J.-J. (1976). *Canada: Inuit Games and Songs*. Paris: Audivis-UNESCO.

Nattiez, J.-J. (1983). The rekkukara of the Ainu (Japan) and the katajjaq of the Inuit (Canada): a comparison. *World Music* 25, 33–44.

Naumov, I. V. (2006). *The History of Siberia*. London: Routledge. doi: 10.4324/9780203027981

Nei, M., and Graur, D. (1984). Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* 17, 73–18. doi: 10.1007/978-1-4615-6974-9_3

Nettl, B. (2005). *The Study of Ethnomusicology: Thirty-one Issues and Concepts*. Champaign, IL: University of Illinois Press.

Nguyen, P. T. (1986). Restructuring the fixed pitches of the Vietnamese dan nguyet lute: a modification necessitated by the modal system. *Asian Music* 18, 56–70. doi: 10.2307/834158

Niemi, J., Lapsui, A., and Lapsui, M. M. (2004). *Network of Songs: Individual Songs of the Ob'Gulf Nenets: Music and Local History as Sung by Maria Maksimovna Lapsui*. Mémoires d. Helsinki: Societe Finno-Ougrienne.

Nikolsky, A. (2015). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.01405

Nikolsky, A. (2016). Evolution of tonal organization in music optimizes neural mechanisms in symbolic encoding of perceptual reality. Part-2: ancient to seventeenth century. *Front. Psychol.* 7:211. doi: 10.3389/fpsyg.2016.00211

Nikolsky, A. (2017). "On the methodology of the analysis of tonal organization of Jaw Harp music [К методам анализа тоновой организации варганной музыки]," in *Systemic Methods of the Research on Musical Culture, International Scientific Practical Conference in Memory of V.V. Mazepus [Системные методы изучения музыкальной культуры, Международная научно-практическая конференция памяти В.В.Мазепуса, 31/X-1/XI 2017*, ed O. V. Novikova (Novosibirsk: Novosibirsk State Conservatory named after Glinka).

Nikolsky, A., Alekseyev, E. Y., Alekseev, I. Y., and Dyakonova, V. (2017). "Prolegomena of modal organization of Jaw harp music on the example of the articulatory degrees of "talking khomus" [Пролегомена ладовой организации варганной музыки на примере использования артикуляционных ступеней в "говорящем" якутском хомусе]," in *Systemic Methods of the Research on Musical culture, International Scientific Practical Conference in Memory of V.V. Mazepus [Системные методы изучения музыкальной культуры, Международная научно-практическая конференция памяти В.В.Мазепуса, 31/X-1/XI 2017*, ed O. V. Novikova (Novosibirsk: Novosibirsk State Conservatory named after Glinka).

Nissen, T., and Haggag, A. (1988). Parasuicidal snow-wandering in Arctic Northern Norway. *Acta Psychiatr. Scand.* 77, 309–313. doi: 10.1111/j.1600-0447.1988.tb05126.x

Nobuhiko, C. (2008). "The music of the Ainu," in *The Ashgate Research Companion to Japanese Music*, eds. A. M. Tokita and D. W. Hughes (Burlington, VT: Ashgate Publishing, Ltd.), 323–344.

Norman-Haignere, S., Kanwisher, N., and McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J. Neurosci.* 33, 19451–19469. doi: 10.1523/JNEUROSCI.2880-13.2013

Novakovsky, S. (1924). Arctic or siberian hysteria as a reflex of the geographic environment. *Ecology* 5, 113–127. doi: 10.2307/1929009

Novik, Y. S. (1998). "Thing-sign" and "thing-gesture": on semiotic interpretation of fetishism ["Вещь-знак" и "вещь-жест": к семиотической интерпретации фетишей]. *Bull. Russ. State Univers. Hum.* 3, 79–97.

Novik, Y. S. (1999). "Semiotic functions of human voice in folklore and beliefs of Siberian peoples [Семиотические функции голоса в фольклоре и

верованиях народов Сибири]," in *Eastern Folklore and Mythology from the Comparative-Typological Aspect*, eds N. Lidova and N. Nikulin (Moscow: Russian Academy of Science), 217–235.

Novik, Y. S. (2003). "Mythological representations of voice in folklore and beliefs of Siberian peoples [Мифологические представления о голосе в фольклоре и верованиях народов Сибири]," in *Eurasian Space: Sound, Word, Image [Евразийское пространство: Звук, слово, образ]*, ed V. V. Ivanov (Moscow: Languages of Slavic Cultures), 134–156.

Novik, Y. S. (2004). *Rite and Folklore in Siberian Shamanism: An Experiment in Correlation of Structures [Обряд и фольклор в сибирском шаманизме: Опыт сопоставления структур]*. Moscow: Eastern Literature, Russian Academy of Science [Восточная литература РАН].

Oakes, J. (1991). Environmental factors influencing bird-skin clothing production. *Arctic Alpine Res.* 23, 71–79. doi: 10.2307/1551440

O'Brien, W. (2014). *Prehistoric Copper Mining in Europe: 5500-500 BC*. Oxford, New York, NY: Oxford University Press.

Ogotoyev, P. P. (1988). "Problems of notation of Yakut khomus [Проблема нотации якутского хомуса]," in *Musical ethnography of North Asia [Музыкальная этнография Северной Азии]*, ed Y. Sheikin (Novosibirsk: Novosibirsk State Conservatory named after Glinka), 150–154.

Ojamaa, T. (2002). The story of life in music: autobiographical songs of the nganasans. *Folklore Electron. J. Folklore* 20–22, 77–97. doi: 10.7592/FEJF2002.21.songs

Ojamaa, T. (2003). "An attempt to segment a nenets melody: inside and outside view," in *Proceedings of the 5th Triennal Conference for the European Society of the Cognitive Sciences of Music (ESCOM): 5th Triennal Conference for the European Society of the Cognitive Sciences of Music (ESCOM)*, eds R. Kopiez, A. C. Lehmann, I. Wolther, and C. Wolf (Hanover: Hanover University of Music and Drama), 527–530.

Ojamaa, T. (2005). Throat rasping: problems of visualization. *World Mus.* 47, 55–69. doi: 10.1080/04419057.2005.9674418

Ojamaa, T., and Ross, J. (2004). "Relationship between texts and tunes in the Siberian folksongs," in *CIM04: Conference on Interdisciplinary Musicology*, eds R. Parncutt, A. Kessler, and F. Zimmer (Graz: Graz University), 134–135.

Ojamaa, T., and Ross, J. (2011). The perceived structure of Forest Nenets songs: a cross-cultural case study. *Psychomusicol. Music Mind Brain* 21, 159–175. doi: 10.1037/h0094010

Okladnikov, A. P. (1968). *Representation of Faces of Ancient Amur [Лики древнего Амура]*. Novosibirsk: West Siberian Book Publishing.

Okladnikov, A. P. (1971). *Petroglyphs of Lower Amur [Петроглифы Нижнего Амура]*. Leningrad: Nauka.

Okladnikov, A. P. (1974). *Petroglyphs of Baikal - Monuments of the Ancient Culture of Siberian Peoples [Петроглифы Байкала – памятники древней культуры народов Сибири]*. Novosibirsk: Nauka.

Okladnikov, A. P. (1979). "Siberian archaeology at present [Сибирская археология на современном этапе]," in *The New in Archaeology of Siberia and Far East [Новое в археологии Сибири и Дальнего Востока]*, ed A. P. Pogozheva (Novosibirsk: Nauka), 5–29.

Okladnikov, A. P., and Mazin, A. I. (1976). *Petroglyphs of Olyokma River and Upper Amur Region [Писаницы реки Олёкмы и Верхнего Приамурья]*. Novosibirsk: Nauka.

Okladnikov, A. P., and Mazin, A. I. (1979). *Petroglyphs of the Aldan River Valley [Писаницы бассеина реки Алдан]*. Novosibirsk: Nauka.

Okladnikov, A. P., and Zaporozhskaya, V. D. (1972). *Petroglyphs of Middle Lena [Петроглифы Средней Лены]*. Leningrad: Nauka.

Oleszczak, Ł., Michalczewski, K., Borodovskii, A. P., and Pokutta, D. A. (2018). Chultukov Log 9 – A settlement from the Xiongnu-Xianbei-Rouran period in the Northern Altai. *Eurasian Prehistory* 14, 153–178.

Pakendorf, B. (2007). *Contact in the Prehistory of the Sakha (Yakuts): Linguistic and Genetic Perspectives*. Available online at: https://openaccess.leidenuniv.nl/handle/1887/12492 (accessed July 14, 2018).

Papoušek, H. (1996). "Musicality in infancy research: biological and cultural origins of early musicality," in *Musical Beginnings: Origins and Development of Musical competence*, ed I. Deliège (New York, NY: Oxford University Press), 37–55. doi: 10.1093/acprof:oso/9780198523321.003.0002

Papoušek, H., and Papoušek, M. (1995). "Beginning of human musicality," in *Music and the Mind Machine: The Psychophysiology and Psychopathology of the Sense of Music* (Berlin; Heidelberg: Springer), 27–34. doi: 10.1007/978-3-642-79327-1_3

Park, J. S., and Gordon, R. B. (2007). Traditions and transitions in Korean bronze technology. *J. Archaeol. Sci.* 34, 1991–2002. doi: 10.1016/j.jas.2007.01.010

Pashina, O. A., Vasilyeva, Y. Y., Danchenkova, N. Y., Dorokhova, Y. A., Lapina, V. A., and Matsiyevsky, I. V. (2005). *Folk Musical Creativity [Народное музыкальное творчество]*. Sankt-Petersburg: Kompozitor [Композитор].

Pashkovskii, O. (2012). Jew's Harps XIII–XVII centuries in archaeological finds [Дримби XIII – Початку XVII Ст. За Археологічними Знахідками]. *New Res. Monum. Cossack Era Ukraine [Нові Дослідження Пам'яток Козацької Доби в Україні]* 12, 248–253.

Patel, A. D. (2010). *Music, Language, and the Brain*. Oxford; New York, NY: Oxford University Press.

Pegg, C. (2001). *Mongolian Music, Dance and Oral Narrative: Performing Diverse Identities*. Seattle, WA: University of Washington Press.

Pekarsky, E. K. (1959). *The Dictionary of Yakut Language [Словарь якутского языка]*. Moscow: Academy of Science of USSR [Изд-во Академии наук СССР].

Petrov, P. A. (1960). *On So-called Polar Hysteria [О так называемой полярной истерии]*. The Collection of Scientific Works of the Yakut Republican Hospital.

Picken, L. (1957). "The music of far eastern Asia," in *New Oxford History of Music: Ancient and Oriental Music*, ed E. Wellesz (London; New York, NY: Oxford University Press), 135–189.

Pignocchi, J. L. (2002). Los Mapuche y el trompe [The Mapuche and the Trompe]. *Vierundzwanzigsteljahrsschrift der International Maultrommelvirtuosengenossenschaft* 10, 31–38.

Pignocchi, J. L. (2004). Trompas in the Zapata Gollan Collection, Found in the Ruins of Santa Fe la Vieja, from the years 1573-1660. *J. Int. Jew's Harp Soc.* 2, 11–21.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 976–986. doi: 10.1037/0096-1523.20.5.976

Plack, C. J. (2010). Musical consonance: the importance of harmonicity. *Curr. Biol.* 20, R476–R478. doi: 10.1016/j.cub.2010.03.044

Popov, A. I. (1949). "Data on history of Yakut religion of the former Viliui District [Материалы по истории религии якутов бывшего Вилюйского округа]," in *The Anthology of the Museum of Anthropology and Ethnography [Сборник музея антропологии и этнографии]*, ed S. P. Tolstov (Moscow; Leningrad: Academy of Science of USSR [Изд-во Академии наук СССР]), 255–323.

Poss, N. F. (2012). *Hmong Music and Language Cognition: An Interdisciplinary Investigation*. Available at: https://oatd.org/oatd/record?record=oai%5C%3Aetd.ohiolink.edu%5C%3Aosu1332472729 (accessed March 23, 2017).

Potanin, G. N. (1883). *The Essays on North-Western Mongolia [Очерки Северо-Западнои Монголии]*. Sankt-Petersburg: V. Kirshbaum Publishing.

Potts, D. T. (2012). "Technological transfer and innovation in Ancient Eurasia," in *The Globalization of Knowledge in History*, ed J. Renn (Berlin: Max Planck Research Library for the History and Development of Knowledge), 105–120. doi: 10.12946/rg22/052-060

Powell, A., Shennan, S., and Thomas, M. G. (2009). Late pleistocene demography and the appearance of modern human behavior. *Science* 324, 1298–1301. doi: 10.1126/science.1170165

Prakash, B., and Tripathi, V. (1986). Iron technology in ancient India. *Hist. Metall.* 568–579.

Pugh-Kitingan, J. (1977). Huli language and instrumental performance. *Ethnomusicology* 21, 205–232. doi: 10.2307/850944

Qin, P., and Stoneking, M. (2015). Denisovan ancestry in east Eurasian and native American populations. *Mol. Biol. Evol.* 32, 2665–2674. doi: 10.1093/molbev/msv141

Radivojević, M., Rehren, T., Pernicka, E., Šljivar, D., Brauns, M., and Borić, D. (2010). On the origins of extractive metallurgy: new evidence from Europe. *J. Archaeol. Sci.* 37, 2775–2787. doi: 10.1016/j.jas.2010.06.012

Rault, L., and Brenton, J. (2000). *Musical Instruments: Craftsmanship and Traditions from Prehistory to the Present*. New York, NY: Harry N. Abrams.

Reedy, C. L. (1997). *Himalayan Bronzes: Technology, Style, and Choices*. Newark: University of Delaware Press.

Reich, D., Reinhard, J., Smith, C., Nordenfelt, S., Richards, S. M., Rohland, N., et al. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Sci. Adv.* 2:e1501385. doi: 10.1126/sciadv.1501385

Relethford, J. H. (1998). Genetics of modern human origins and diversity. *Annu. Rev. Anthropol.* 27, 1–23. doi: 10.1146/annurev.anthro.27.1.1

Repina, T. A. (1966a). "On the problem of the mechanisms of objectivitization of child's pitch distinctions [К вопросу о механизмах явления «опредмечивания» в звуковысотном различении ребенка]," in *Development of Perception in Early and Preschool Childhood. [Развитие восприятия в раннем и дошкольном детстве]*, eds A. V Zaporozhets and M. I. Lisina (Moscow: Prosvesheniye), 98–141.

Repina, T. A. (1966b). "Perception of pitch differences in relation to organization of activitiy of preschool age children [Восприятие звуковысотных различий в зависимости от организации деятельности детей дошкольного возраста]," in *Development of Perception in Early and Preschool Childhood [Развитие восприятия в раннем и дошкольном детстве]*, eds A. V Zaporozhets and M. I. Lisina (Moscow: Prosvesheniye), 74–97.

Réti, R. (1951). *The Thematic Process in Music*. New York, NY: Macmillan Publishers.

Robb, J. (1993). A social prehistory of European languages. *Antiquity* 67, 747–760. doi: 10.1017/S0003598X00063766

Robbeets, M. (2015). *Diachrony of Verb Morphology: Japanese and the Transeurasian Languages*. Berlin: De Gruyter Mouton.

Roberts, B. W., Thornton, C. P., and Pigott, V. C. (2009). Development of metallurgy in Eurasia. *Antiquity* 83, 1012–1022. doi: 10.1017/S0003598X00099312

Robertson-DeCarbo, C. E. (1974). Music as therapy: a bio-cultural problem. *Ethnomusicology* 18, 31. doi: 10.2307/850058

Robinson, E. (1942). *Shell Fishhooks of the California Coast*. Available online at: https://scholar.google.com/scholar?hl=en&as_sdt=2005&sciodt=0%2C5&cites=10601864673987735240&scipsc=&q=%22Shell$+$fishhooks$+$of$+$the$+$California$+$coast%22&btnG= (accessed September 3, 2019).

Rogers, A. R. (1997). "Population structure and modern human origins," in *Progress in Population Genetics and Human Evolution*, eds P. J. Donnelly and S. Tavare (New York, NY: Springer), 55–79. doi: 10.1007/978-1-4757-2609-1_4

Rogers, A. R., and Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* 9, 552–569.

Rouget, G. (1985). *Music and Trance: A Theory of the Relations Between Music and Possession*. Chicago, IL: University of Chicago Press.

Rubinson, K. (1992). "A reconsideration of pazyryk," in *Foundations of Empire: Archaeology and Art of the Eurasian Steppes,* ed G. Seaman (Los Angeles, CA: Ethnographic Press, University of Southern California), 68–76.

Russo, F. A., and Thompson, W. F. (2005). An interval size illusion: the influence of timbre on the perceived size of melodic intervals. *Percept. Psychophys.* 67, 559–568. doi: 10.3758/BF03193514

Rybakov, B. A. (1987). *Archaeology of the USSR. Bronze Age of the Forest Zone of the USSR [Археология СССР. Эпоха бронзы лесной полосы СССР]*. Moscow: Nauka.

Sachs, C. (1917). Die maultrommel. Eine typologische vorstudie. *Zeitschrift für Ethnologie* 49, 185–200.

Sachs, C. (1940). *The History of Musical Instruments*. New York, NY: Dover Publications.

Sagalayev, A. M., and Oktiabr'skaya, I. V. (1990). *Traditional Worldview of Turks of Southern Siberia. Sign and ritual. Традиционное мировоззрение тюрков Южнои Сибири. Знак и ритуал]*. Novosibirsk: Nauka.

Saitis, C., and Weinzierl, S. (2019). "The semantics of timbre," in *Timbre: Acoustics, Perception, and Cognition,* eds K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, and R. R. Fay (Cham: Springer), 119–150. doi: 10.1007/978-3-030-14832-4_5

Sakaki, Y. (1903). Über das Imubacco, eine dem Jumping, Latnh und Meriachenje nahe verwandte Neurose des Ainovolkes. *Neurologia* 1, 7.

Sam, S.-A. (2008). "The khmer people of Cambodia," in *The Garland Handbook of Southeast Asian Music*, eds T. Miller and S. Williams (London: Routledge), 85–120.

Sammler, D., Baird, A., Valabrègue, R., Clément, S., Dupont, S., Belin, P., et al. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *J. Neurosci.* 30, 3572–3578. doi: 10.1523/JNEUROSCI.2751-09.2010

Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining "blended" instrument pairings in orchestration. *Music Percept. Interdiscipl. J.* 13, 209–246. doi: 10.2307/40285694

Sarangerel (2001). *Chosen by the Spirits: Following Your Shamanic Calling*. Merrimac, MA: Destiny Books.

Schneider, A. (2001). Sound, pitch, and scale: from "tone measurements" to sonological analysis in ethnomusicology. *Ethnomusicology* 45, 489–519. doi: 10.2307/852868

Schouten, J. F. (1940). "The residue and the mechanism of hearing," in *Proceedings of the Koninklijke Nederlandsche Akademie von Wetenschappen*. Available online at: https://www.boekwinkeltjes.nl/b/172949494/The_residue_and_the_mechanism_of_hearing/ (accessed December 30, 2017).

Schouten, J. F., Ritsma, R. J., and Cardozo, B. L. (1962). Pitch of the residue. *J. Acous. Soc. Am.* 34, 1418–1424. doi: 10.1121/1.1918360

Schurr, T. G. (2015). "Tracing human movements from siberia to the Americas: insights from genetic studies," in *Mobility and Ancient Society in Asia and the Americas*, eds M. D. Frachetti and R. N. I. Spengler (Cham: Springer International Publishing), 23–47. doi: 10.1007/978-3-319-15138-0_4

Schwartz, D. A., Howe, C. Q., and Purves, D. (2003). The statistical structure of human speech sounds predicts musical universals. *J. Neurosci.* 23, 7160–7168. doi: 10.1523/JNEUROSCI.23-18-07160.2003

Schwarzer, G. (1997). Analytic and holistic modes in the development of melody perception. *Psychol. Mus.* 25, 35–56. doi: 10.1177/0305735697251004

Scruton, R. (1997). *The Aesthetics of Music*. New York, NY: Clarendon Press.

Seliutina, I. Y., Urtegeshev, A. Y., Letiagin, A. Y., Shevela, A. I., Dobrinina, A. A., Esenbayeva, G. A., et al. (2012). *Articulatory Databases of Indigenous Turkic Ethnoses of Southern Siberia (According to Magnetic Resonance Tomography and Digital Rentgenography) [Артикуляторные базы коренных тюркских этносов Южной Сибири (по данным МРТ и цифровой рентгенографии)]*. Novosibirsk: Russian Academy of Science, Siberian Department.

Sergeyeva, N. F. (1981). *The Oldest Copper Metallurgy in South Eastern Siberia [Древнейшая металлургия меди юга Восточной Сибири]*. Moscow: Nauka.

Sermier, C. (2002). *Mongolia: Empire of the Steppes*. Fyshwick: Odyssey.

Shamma, S., and Klein, D. (2000). The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *J. Acous. Soc. Am.* 107, 2631–2644. doi: 10.1121/1.428649

Shchurov, V. (1995). *Khomus: Jew's Harp Music of Turkic Peoples in the Urals, Siberia, and Central Asia.* Leiden: PAN Records.

Sheikin, Y. I. (1982). *Problem of Genre in Musical Folklore of Udege [Проблема Жанра в Музыкальном Фольклоре Удэ]* (Dissertation). The Department of Music, State Institute of Theater, Music and Cinematography. Leningrad, Russia.

Sheikin, Y. I. (1986a). "Musical instruments of Jurchen [Музыкальные инструменты чжурчженеи]," in *Problems of Traditional Instrumental Music of Peoples of the USSR: Instrument-Performer-Music [Проблемы традиционной инструментальной музыки народов СССР: Инструмент-исполнитель-музыка]*, ed I. Matsiyevskii (Leningrad: LGITMiK), 100–109.

Sheikin, Y. I. (1986b). "Musical instruments of Ude: etimology, construction, playing patterns [Музыкальные инструменты удэ: этимология, конструкции, наигрыши]," in *Musical creativity of the peoples of Siberia and Far East [Музыкальное творчество народов Сибири и Дальнего Востока]*, ed I. I. Golovneva (Novosibirsk: Novosibirsk State Conservatory named after Glinka), 38–72.

Sheikin, Y. I. (1990). *Instrumental Music of Yugra [Инструментальная музыка Югры]*. Novosibirsk: Novosibirsk State Conservatory named after Glinka.

Sheikin, Y. I. (1996). *Musical Culture of Peoples of Northern Asia [Музыкальная культура народов Северной Азии]*. Yakutsk: Yakutskii Scientific Center.

Sheikin, Y. I. (2002). *History of Music Culture of Siberian Ethnicities: A Comparative Historic Investigation [История музыкальной культуры народов Сибири: сравнительно-историческоеисследование]*. Moscow: Eastern Literature, Russian Academy of Science Восточная литература РАН].

Sheikin, Y. I. (2018). *Musical Culture of Chukchi [Музыкальная культура чукчей]*, eds O. Dobzhanskaya and T. Ignatyeva Novosibirsk. Russia: Nauka.

Shennan, S. (2001). Demography and cultural innovation: a model and its implications for the emergence of modern human culture. *Cambridge Archaeol. J.* 11, 5–16. doi: 10.1017/S0959774301000014

Sherry, S. T., Rogers, A. R., Harpending, H., Soodyall, H., Jenkins, T., and Stoneking, M. (1994). Mismatch distributions of mtDNA reveal recent human population expansions. *Hum. Biol.* 66, 761–755.

Shimkevich, P. P. (1896). *Materials for the Study of Shamanism Amongst Goldi [Материалы для изучения шаманства у гольдов]*, ed M. Y. Sibirtsev. Khabarovsk: The Notes of the Imperial Russian Geogrphic Society, Priamurye Department.

Shimomura, I. (2016). "Acoustic mechanism of sound generation of the "talking Jaw harp" and of the game of "larynx substitution". [Акустический механизм звукоизвлечения из «говорящего» варгана и из игры «замена горлом»]," in *Folklore of Paleoasiatic Peoples: Proceedings of the II International Scientific Conference [Фольклор палеоазиатских народов: материалы II Международной научной конференции, г. Якутск, 21–25 ноября 2016 г.]* ed A. N. Alekseyev (Yakutsk: RIO Media Holding), 216–221.

Shinohara, K., and Kawahara, S. (2010). A cross-linguistic study of sound symbolism: the images of size. *Annu. Meet. Berkeley Linguist. Soc.* 36, 396. doi: 10.3765/bls.v36i1.3926

Shirokogoroff, S. M. (1929). *Phychomental Complex of the Northern Tungus*. Shanghai: Garland Pub.

Shishigin, S. S. (1995). *Play Khomus [Играйте на хомусе]*. Yakutsk: Ministry of Culture of Sakha Republic.

Shishigin, S. S. (2015). *Learn to Play Khomus [Хомус тарда үөрэнин]*. Yakutsk: RIO Media Holding.

Shreibler, S. I. (1927). *The Preliminary Report of the Medical Sanitary Team of the Yakut Expedition of the Academy of Science of the USSR 1925-26 on the inspection of the Viliui and Olekma districts [Предварительный отчёт медико-санитарного Отряда Якутской Экспедиции Академии На.]*. Leningrad: Academy of Science of USSR.

Shrenk, L. I. (1903). *On Aliens of the Amur Region. Ethnographic Section 2: Main Traits of Family, Social and Personal Life [Об инородцах Амурского края. Этнографическая часть. Вторая половина. Основные черты семейной, общественной и внутренней жизни]*. Sankt-Petersburg: The Imperial Academy of Science of Russia.

Shternberg, L. Y. (1936). *Prehistoric Religion in Light of Ethnography: Research, Articles, Lectures [Первобытная религия в свете этнографии: исследования, статьи, лекции]*. Leningrad: Institute of Peoples of the North at the Central Executive Committee of the USSR).

Shulga, P. I. (2015). *The Yuhuangmiao Cemetery in Northern China (the 7th – 6th centuries B.C.) [Могильник Юйхуанмяо в Северном Китае (VII-VI века до нашей эры)]*. Novosibirsk: Publishing department of the institute of archaeology and ethnography, Russian Academy of Science.

Sidéra, I. (2011). Fabriquer des cuillers en os: L'exemple de Kovacevo. *Stud. Praehistor.* 14, 55–62.

Sieroszewski, W. (1896). *The Yakuts. An Ethnographic Study [Якуты. Опыт этнографического исследования]*. Sankt-Petersburg: The Imperial Geographic Society Publishing.

Sikora, M., Pitulko, V. V., Sousa, V. C., Allentoft, M. E., Vinner, L., Rasmussen, S., et al. (2019). The population history of northeastern Siberia since the Pleistocene. *Nature* 570, 182–188. doi: 10.1038/s41586-019-1279-z

Simana, S., and Preisig, E. (2003). "Kmhmu' Music and Musical Instruments," in *Laosand Ethnic Minority Cultures: Promoting Heritage*, ed Y. Goudineau (Paris: UNESCO Publishing), 123–132.

Simons, G. (1986). Early childhood musical development: a survey of selected research. *Bull. Council Res. Mus. Educ.* 86, 36–52.

Simukhin, A. I. (2006). *Technological traditions in production of nonferrous metallic artifacts in transbaikalia [Технологические Традиции в Производстве Изделий Из Цветного Металла в Забайкалье]* (Dissertation). State Buryat University, Russian Academy of Science, Ulan-Ude, Russia, The Department of Ethnography, Ethnology and Anthropology.

Singh, P. G., and Hirsh, I. J. (1992). Influence of spectral locus and F0 changes on the pitch and timbre of complex tones. *J. Acous. Soc. Am.* 92, 2650–2661. doi: 10.1121/1.404381

Skrebkova-Filatova, M. S. (1985). *Texture in Music [Фактура в музыке]*. Moscow: Muzyka [Музыка].

Slobin, M. (1976). *Music in the Culture of Northern Afghanistan. Viking Fun*. Tucson, AZ: Wenner-Gren Foundation for Anthropological Research.

Small, C. (1998). *Musicking: The Meanings of Performing and Listening*. Hanover, NH: University Press of New England.

Soares, P., Alshamali, F., Pereira, J. B., Fernandes, V., Silva, N. M., Afonso, C., et al. (2012). The Expansion of mtDNA Haplogroup L3 within and out of Africa. *Mol. Biol. Evol.* 29, 915–927. doi: 10.1093/molbev/msr245

Solomonova, N. V. (2000). *Musical culture of peoples of the Russian Far East of the 19th–20th centuries: Thnomusicological Essays [Музыкальная Культура Народов Дальнего Востока России XIX-XX Вв. (Этномузыкологические Очерки)]* (Dissertation). The Department of Music Theory, The Moscow State Conservatory named after Tchaikovsky.

Starostin, S. A. (1991). *The Altaic Problem and Origins of Japanese Language [Алтаиская проблема и происхождение японского языка]*. Moscow: Nauka.

Startsev, A. F. (2017). *Ethnic Beliefs of Tungus-Manchurians Regarding Nature and Society [Этнические представления тунгусо-маньчжуров о природе и обществе]*. Vladivostok: Dal'nauka.

Stépanoff, C., Marchina, C., Fossier, C., and Bureau, N. (2017). Animal autonomy and intermittent coexistences: north Asian modes of herding. *Curr. Anthropol.* 58, 57–81. doi: 10.1086/690120

Stige, B. (2002). *Culture-Centered Music Therapy*. Barcelona: Barcelona Publishers.

Stockmann, E. (1987). "Researching European folk musical instruments and describing them in a multivolume reference book [Исследование народных музыкальных инструментов Европы и их описание в многотомном справочнике (Handbuch)]," in *Folk Musical Instruments and Instrumental Music [Народные музыкальные инструменты и инструментальная музыка]*, ed I. V. Matsiyevskii (Moscow: Soviet Composer [Советский композитор]), 39–56.

Stoter, F.-R., Schoeffler, M., Edler, B., and Herre, J. (2013). Human ability of counting the number of instruments in polyphonic music. *J. Acous. Soc. Am.* 133, 3366–3366. doi: 10.1121/1.4805760

Subbotina, A. (2008). *Monuments of early iron age of Chundo type in Korean Peninsula [Памятники Раннего Железного Века Типа Чундо На Корейском Полуострове]* (Dissertation). The Department of Archaeology, Russian Academy of Science, Siberian Section, Novosibirsk, Russia.

Suleimanova, R. N. (1994). "The remnants of shamanism amongst Bashkirs [Пережитки шаманства у башкир]," in *The Ethnological Research in Bashkortostan [Этнологические исследования в Башкортостане]*, eds I. Gabdrafikov, M. D. Kiyekbayev, and F. A. Shakurova (Ufa: Russian Academy of Science), 119–130.

Sunchugashev, Y. I. (1975). *The Oldest Mines and Monuments of Early Metallurgy in Khakass-Minusinsk Basin [Древнейшие рудники и памятники ранней металлургии в Хакасско- Минусинской котловине]*. Moscow: Nauka.

Sundberg, J. (1974). Articulatory interpretation of the "singing formant." *J. Acous. Soc. Am.* 55, 838–844. doi: 10.1121/1.1914609

Sundberg, J. (1987). *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press.

Suzukei, V. (1989). *Tuvan Traditional Musical Instruments [Тувинские традиционные музыкальные инструменты]*. Kyzyl: Tuvan Book Publishing [Тувинское книжное издательство].

Suzukei, V. (1991). "The typology of Tuvan khomuses [Типология тувинских хомусов]," in *Jew's Harp (khomus) and Its Music. Proceedings of the First All-Union conference, 1988* [Варган (хомус) и его музыка. Материалы I Всесоюзной конференции 1988], ed E. Y. Alekseyev (Yakutsk: Yakut Institute of language, literature and history of the USSR Academy of Science), 59–67.

Suzukei, V. (2006). *The Configuration of the Development of Tuvan Musical Culture: Dynamics of the Axiological Aspect [Конфигурация развития*

*музыкальной культуры Тувы: динамика аксиологического аспекта]*. Kemerovo: Kemerovo University of Culture and Arts.

Suzukei, V. (2010). *Khomus in the traditional culture of the Tuvans [Хомус в традиционной культуре тувинцев]*. Kyzyl: Tyvapoligraf.

Svantesson, J. O. (2017). Sound symbolism: the role of word sound in meaning. *Wiley Interdiscipl. Rev. Cogn. Sci.* 8:e01441. doi: 10.1002/wcs.1441

Tadagawa, L. (1991). On Japanese Jaw Harps of the millenial age [О японских варганах тысячелетней давности]. *Khomus* 1, 24–47.

Tadagawa, L. (2001). *Mukkuri Hawehe: Sound of Mukkuri. Jew's Harp and Vocal Music of the Ainu*. Saitama: Nihon Koukin Kyoukai.

Tadagawa, L. (2007). Asian excavated Jew's harps: a checklist. *J. Int. Jews Harp Soc.* 4, 5–11.

Tadagawa, L. (2016). Asian excavated Jew's Harps: a checklist (1) - lamellate Jew's harps (1). *Instit. Ethnomusicol. Bull. Tokyo Coll. Mus.* 5, 57–70.

Tadagawa, L. (2017a). Asian excavated Jew's Harps: a checklist (2) - lamellate jew's harps (2). *Instit. Ethnomusicol. Bull. Tokyo Coll. Mus.* 6, 57–68.

Tadagawa, L. (2017b). The khomus is my red deer on which I fly through the middle world (Khomus in the shamanic practice of Tuva: Research issues). *New Res. Tuva* 2, 165–176. doi: 10.25178/nit.2017.2.7

Tadysheva, N. O. (2016). Traditional etiquette of Altaians in the aspect of family relations [Традиционный этикет алтайцев в семейно-родственном аспекте]. *Tomsk J. Linguist. Anthropol. Res.* 2, 98–105.

Tadysheva, N. O. (2018). The reflection of tree cult in traditional beliefs of Sayano-Altaic Turks [Отражение культа дерева в традиционном мировоззрении тюрков Саяно-Алтая]. *Tomsk State Univers. J. History* 51, 118–123. doi: 10.17223/19988613/51/17

Tallavaara, M., Eronen, J. T., and Luoto, M. (2018). Productivity, biodiversity, and pathogens influence the global hunter-gatherer population density. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1232–1237. doi: 10.1073/pnas.1715638115

Tarasov, P. E., Kvavadze, E. V., Webb, T., Dorofeyuk, N. I., Bykova, G. V., Volkova, V. S., et al. (2000). Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. *J. Biogeogr.* 27, 609–620. doi: 10.1046/j.1365-2699.2000.00429.x

Tardieu, D., and McAdams, S. (2012). Perception of dyads of impulsive and sustained instrument sounds. *Music Percept. Interdiscipl. J.* 30, 117–128. doi: 10.1525/mp.2012.30.2.117

Tchakhov, A. I. (2012). *Ancient Yakut Musical Instruments: The Manufacturing Technology [Старинные якутские музыкальные инструменты: технология изготовления]*. Yakutsk: Alaas.

Tchernetsov, V. N. (1971). *Rock Images of Ural [Наскальные изображения Урала]*. Moscow: Nauka.

Tchernetsov, V. N. (1987). *Sources on Ethnography of Western Siberia [Источники по этнографии Западной Сибири]*. Tomsk: The State Tomsk University.

Tenishev, E. R. (1984). *Comparative Historic Grammar of Turkic Languages. Phonetics [Сравнительно-историческая грамматика тюркских языков]*. Moscow: Nauka.

Tenishev, E. R. (1997). "Altaic languages [Алтайские языки]," in *Languages of the World. Turkic Languages [Языки мира: Тюркские языки]* (Bishkek: Russian Academy of Science, Kyrkyzstan), 7–16.

Teplov, B. (1947). *The Psychology of Musical Abilities [Психология музыкальных способностей]*. Moscow: Academy of Pedagogical Sciences of Russia.

Terhardt, E. (1974). Pitch, consonance, and harmony. *J. Acous. Soc. Am.* 55, 1061–1069. doi: 10.1121/1.1914648

Terhardt, E., Stoll, G., and Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *J. Acous. Soc. Am.* 71, 679–688. doi: 10.1121/1.387544

Thiel, B. (1989). Excavations at musang cave, Northeast Luzon, Philippines. *Asian Perspect.* 28, 61–81.

Titze, I. R. (1988). A framework for the study of vocal registers. *J. Voice* 2, 183–194. doi: 10.1016/S0892-1997(88)80075-4

Tokarsky, A. A. (1893). *Meriachenye and the Disease of Convulsive Contractions [Меряченье и болезнь судорожных подергиваний]*. Moscow: Vrach.

Tompkins, D. (2010). *How to Wreck a Nice Beach : the Vocoder from World War II to Hip-Hop : The Machine Speaks*. Brooklyn, NY: Melville House.

Trainor, L. J., Wu, L., and Tsang, C. D. (2004). Long-term memory for music: infants remember tempo and timbre. *Dev. Sci.* 7, 289–296. doi: 10.1111/j.1467-7687.2004.00348.x

Trehub, S. E., Endman, M. W., and Thorpe, L. A. (1990). Infants' perception of timbre: classification of complex tones by spectral structure. *J. Exp. Child Psychol.* 49, 300–313. doi: 10.1016/0022-0965(90)90060-L

Treitler, L. (2007). "'Centonate' chant: übles flickwerk or e pluribus unus?" in *With Voice and Pen*, ed L. Treitler (Oxford: Oxford University Press), 186–201. doi: 10.1093/acprof:oso/9780199214761.003.0007

Trevarthen, C. (2008). The musical art of infant conversation: narrating in the time of sympathetic experience, without rational interpretation, before words. *Mus. Sci.* 12, 15–46. doi: 10.1177/1029864908012001021

Trias, S. (2010). *Helmholtz and Coupled Resonators Acoustics in Jew's Harp Playing* (Dissertation). Faculty of Arts, Telemark University College, Notodden, Norway.

Tripathi, V. (2018). Metals and metallurgy in the harappan civilization. *Indian J. History Sci.* 53, 279–295. doi: 10.16943/ijhs/2018/v53i3/49460

Tseng, W.-S. (2003). *Clinician's Guide to Cultural Psychiatry*. San Diego, CA: Academic Press.

Tseng, W.-S. (2007). "Culture and psychopathology: general view," in *Textbook of Cultural Psychiatry*, eds D. Bhugra and K. Bhui (Cambridge, UK: Cambridge University Press), 95–112. doi: 10.1017/CBO9780511543609.010

Turbat, T. (2017). "Archaeological Jew's Harps from Ancient Nomadic Cultures of Eastern Eurasia [Евразийн зүүн хэсгийн эртний нүүдэлчдийн археологийн дурсгалаас илэрсэн хэл хуурууд]," in *Archaeological Survey*, ed T. Turbat (Ulaanbaatar: Academic Council of the Institute of History and Archeology of the Ministry of Nature, Environment and Tourism), 169–180.

Tylecote, R. F. (2002). *A History of Metallurgy*. London: Maney Pub., for the Institute of Materials.

Tyukhteneva, S. P. (2015). Personality and society amongst Altaians: from clan affiliation to all-Altaic identity [Личность и общество у алтаицев: от родовои принадлежности до общеалтаискои идентичности]. *Bull. Kalmyk Instit. Humanitarian Stud. Russ. Acad. Sci.* 4, 72–81.

Uchida, R., and Catlin, A. (2008). "Music of upland minorities in Burma, Laos, and Thailand," in *The Garland Handbook of Southeast Asian Music*, eds T. Miller and S. Williams (London: Routledge), 303–316.

Ukhtomsky, A. A. (1978). *Selected Works [Избранные труды]*. Leningrad: Nauka.

Vainshtein, S. (1991). *The World of Nomads of the Center of Asia [Мир кочевников центра Азии]*. Moscow: Nauka.

Val'kova, V. B. (1992). *Musical Thematicism, Cognition, Culture [Музыкальный тематизм, мышление, культура]*. Nizhnii Novgorod: State Nizhegorodskii University.

van der Schyff, D. (2013). Music, culture and the evolution of the human mind: looking beyond dichotomies. *Hellenic J. Mus. Educ. Cult.* 4, 1–11.

Vasilevich, G. M. (1969). *Evenki: The Historic-Ethnographic Sketches (18th-Beginning of 20th Centuries) [Эвенки: историко-этнографические очерки (XVIII-начало XX в.)]*. Leningrad: Nauka, The Institute of Ethnography of the USSR Academy of Science.

Vasilyev, V. Y. (2016). Genesis of shamanic instruments of Sakha people in light of the image of Mother-Beast of Turks, Mongols and Tungus [Генезис шаманских инструментов народа саха в свете образа матери-зверя у тюрков, монголов и тунгусов]. *Northeastern Humanitarian Bull.* 4, 18–27.

Vensten-Tagrina, Z. V. (2008). *The Phenomenon of Personal Song in Traditional Culture of Chukcha [Феномен личной песни в традиционной культуре чукчей]* (Dissertation). Department of Ethnoculturology, Sankt-Petersburg, Russia.

Verbitsky, V. I. (1865). Seoks (bone, kin, generation) of Altaians [Сеоки (кость, род, поколение) алтайцев]. *Tomsk Provincial Gazette* 42, 14.

Vernot, B., Tucci, S., Kelso, J., Schraiber, J. G., Wolf, A. B., Gittelman, R. M., et al. (2016). Excavating neandertal and denisovan DNA from the genomes of melanesian individuals. *Science* 352, 235–239. doi: 10.1126/science.aad9416

Vertkov, K., Blagodatov, G., and Yazovitskaya, E. (1963). *The Atlas of Musical Instruments of the USSR Peoples [Атлас музыкальных инструментов народов СССР]*. Moscow: Muzgiz.

Vinogradov, V. S. (1938). *The Music of Soviet Kyrghistan [Музыка Советской Киргизии]*. Moscow: The Department of Art Management at the Government of Kyrghyz Republic.

Vinogradov, V. S. (1958). *Kirghiz Folk Music [Киргизская народная музыка]*. Frunze: Kirghizgosizdat [Kirghiz State Publishing].

Vitashevsky, N. A. (1911). From the field of prehistoric psycho-neurosis [Из области первобытного психоневроза]. *Ethnogr. Rev.* 1–2, 180–228.

Voldina, T. V. (2017). "Song as a curative remedy in traditional culture of Yugric people of Ob' [Песня как целительное средство в традиционнои культуре обских угров]," in *Medical Ethnography: Contemporary Approaches and Conceptions [Медицинская этнография: современные подходы и концепции]*, ed V. I. Kharitonova (Moscow: Institute of Ethnography & Anthropology of Russian Academy of Science, OOO Publicity), 141–151.

Walker, R. (1997). Visual metaphors as music notation for sung vowel spectra in different cultures. *J. N. Music Res.* 26, 315–345. doi: 10.1080/09298219708570733

Wallace, A. F. C., and Ackerman, R. E. (1960). An interdisciplinary approach to mental disorder among the polar eskimos of northwest Greenland. *Anthropologica* 2:249. doi: 10.2307/25604472

Wan, X. (2011). *Early Development of Bronze Metallurgy in Eastern Eurasia*. Philadelphia, PA: University of Pennsylvania.

Warrier, C. M., and Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Percept. Psychophys.* 64, 198–207. doi: 10.3758/BF03195786

Welch, G. F. (1994). The assessment of singing. *Psychol. Mus.* 22, 3–19. doi: 10.1177/0305735694221001

White, J. C., and Hamilton, E. G. (2009). The transmission of early bronze technology to thailand: new perspectives. *J. World Prehistory* 22, 357–397. doi: 10.1007/s10963-009-9029-z

Whitridge, P. (2015). The sound of contact: Historic Inuit music-making in northern Labrador. *N. Atlantic Archaeol.* 4, 17–42.

Williams, D. B. (1990). Effects of selected music factors on second- and fifth-grade children's perception of melodic motion. *Psychomusicol. J. Res. Music Cogn.* 9, 59–78. doi: 10.1037/h0094158

Wright, J. (2001). "Jew's harp," in *The New Grove Dictionary of Music and Musicians*, eds S. Sadie and J. Tyrrell (London: Macmillan Publishers). doi: 10.1093/gmo/9781561592630.article.14300

Wright, M. (2004). The search for the origins of the Jew's harp. *Silk Road* 2, 49–55.

Wright, M. (2011). The Jew's Harp Trade in Colonial America. *Galpin Soc. J.* 64, 209–218. Available online at: https://www.jstor.org/stable/23209396

Wright, M. (2015). *The Jews-Harp in Britain and Ireland*. Abingdon; Oxfordshire: Routledge.

Wright, M., and Impey, A. (2007). The Birmingham-KwaZulu connection. *J. Int. Jews Harp Soc.* 4, 44–48. doi: 10.1556/SMus.48.2007.3-4.3

Yakovlev, V. I. (2001). *Traditional musical instruments of Volgo-Uralic PEoples: Formation, development and functionality* [Традиционные Музыкальные Инструменты Народов Волго-Уралья: Формирование, Развитие и Функционирование] (Dissertation). The Department of Ethnography, Ethnology and Anthropology, The State Kazan University, Kazan, Russia.

Yanin, V. L. (2006). *Archaeology [Археология]*. Moscow: The Moscow State University.

Yekeyeva, E. V. (2011). Traditions of family upbringing amongst Altaians [Традиции семейного воспитания у алтайцев]. *Courier North-Eastern Fed. Univ.* 8, 103–106.

Yekeyeva, E. V. (2014). "The system of family upbringing [Система семеибного воспитания]," in *The Altaians: Ethnic History. Traditional Culture. Modern Development,* ed N. V. Yekeyev (Gorno-Altaisk: The scientific research institute of Altaistics named after Surazakov), 130–148.

Yenchinov, E. V. (2009). *Family values of Altaians: Transformation of customary law in modern culture [Семейные Ценности Алтайцев: Трансформация Обычного Права в Современной Культуре]* (Dissertation). Moscow State University named after Lomonosov, The Department of Theory and History of Culture.

Yesipova, M. V. (2008)."Phono-instruments (etnophones)," in *Musical Instruments [Музыкальные Инструменты]*, Vol. 633 ed M. V. Yesipova (Moscow: Deka-VS).

Yesipova, M. V., Wright, J., Sheikin Y. I., Gerasimov, O. M., Gorokhovik, E. M., Putilov, B. N., et al. (2008). "Jew's harp [Варган]," in *Musical Instruments [Музыкальные Инструменты]*, ed M. B. Yesipova (Moscow: Deka-VS), 112–115.

Yousof, G.-S. (2004). *The Encyclopedia of Malaysia: Performing Arts*. Brooklyn, NY: Archipelago Press.

Zabolotskaya, P. Y. (2009). *Folkloric Theater of Yakuts: An Attempt of Historic-Hystrionic Research [Шаманские маски народов Сибири)]*. Ulan-Ude: East Siberian State Academy of Culture and Arts.

Zabolotskaya, P. Y. (2011). Shamanic masks of Siberian peoples [Шаманские маски народов Сибири]. *Siberian Philol. J.* 3, 24–34. doi: 10.17223/18137083/36/3

Zagretdinov, R. A. (1997). *The School of Playing Kubyz: A Practical Methodological Aid [Школа игры на кубызе: Учебно-Методическое Пособие]*. Ufa: Belaya Reka.

Zamoruyev, V. (2004). Quaternary glaciation of north-eastern Asia. *Dev. Q. Sci.* 2, 321–323. doi: 10.1016/S1571-0866(04)80137-4

Zaporozhets, A. V. (1985). *Selected Works on Psychology [Избранные Психологические Труды]*, Vol. 1, eds V. V. Davydova and V. P. Zinchenko. Moscow: Pedagogika.

Zaporozhets, A. V. (2003). The development of sensations and perceptions in early and preschool childhood. *J. Russ. East Eur. Psychol.* 40, 22–34. doi: 10.2753/RPO1061-0405400322

Zarate, J. M., Ritson, C. R., and Poeppel, D. (2013). The effect of instrumental timbre on interval discrimination. *PLoS ONE* 8:e75410. doi: 10.1371/journal.pone.0075410

Zemtsovsky, I. (1987). "Musical instrument and musical cognition (on presentation of a problem) [Музыкальный инструмент и музыкальное мышление (к постановке вопроса)]," in *Folk Musical Instruments and Instrumental Music [Народные музыкальные инструменты и инструментальная музыка]*, ed I. V. Matsiyevskii (Moscow: Soviet Composer [Советский композитор]), 125–131.

Zhilich, S., Rudaya, N., Krivonogov, S., Nazarova, L., and Pozdnyakov, D. (2017). Environmental dynamics of the Baraba forest-steppe (Siberia) over the last 8000 years and their impact on the types of economic life of the population. *Q. Sci. Rev.* 163, 152–161. doi: 10.1016/j.quascirev.2017.03.022

Zhirkova, R. R. (1991). *The Methodic Supplement for Learning to Play Khomus [Методическое пособие для разучивания игры на хомусе]*. Yakutsk: Yakutsk University.

Zhuravlev, A. P. (1977). *Eneolithic Karelia [Энеолит Карелии]* (Dissertation). The Department of Archaeology, The Institute of Language, Literature and History of Karelian chapter of the USSR Academy of Sciences, Petrozavodsk, Russia.

Zsiga, E. C. (2013). *The Sounds of Language: an Introduction to Phonetics and Phonology*. Oxford UK: Wiley-Blackwell.

**Conflict of Interest:** AN was employed by company, Braavo Enterprises.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Nikolsky, Alekseyev, Alekseev and Dyakonova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 23 July 2020
doi: 10.3389/fpsyg.2020.01358

# The Pastoral Origin of Semiotically Functional Tonal Organization of Music

*Aleksey Nikolsky**

*Independent Researcher, Austin, TX, United States*

This paper presents a new line of inquiry into when and how music as a semiotic system was born. Eleven principal expressive aspects of music each contains specific structural patterns whose configuration signifies a certain affective state. This distinguishes the tonal organization of music from the phonetic and prosodic organization of natural languages and animal communication. The question of music's origin can therefore be answered by establishing the point in human history at which all eleven expressive aspects might have been abstracted from the instinct-driven primate calls and used to express human psycho-emotional states. Etic analysis of acoustic parameters is the prime means of cross-examination of the typical patterns of expression of the basic emotions in human music versus animal vocal communication. A new method of such analysis is proposed here. Formation of such expressive aspects as meter, tempo, melodic intervals, and articulation can be explained by the influence of bipedal locomotion, breathing cycle, and heartbeat, long before Homo sapiens. However, two aspects, rhythm and melodic contour, most crucial for music as we know it, lack proxies in the Paleolithic lifestyle. The available ethnographic and developmental data leads one to believe that rhythmic and directional patterns of melody became involved in conveying emotion-related information in the process of frequent switching from one call-type to another within the limited repertory of calls. Such calls are usually adopted for the ongoing caretaking of human youngsters and domestic animals. The efficacy of rhythm and pitch contour in affective communication must have been spontaneously discovered in new important cultural activities. The most likely scenario for music to have become fully semiotically functional and to have spread wide enough to avoid extinctions is the formation of cross-specific communication between humans and domesticated animals during the Neolithic demographic explosion and the subsequent cultural revolution. Changes in distance during such communication must have promoted the integration between different expressive aspects and generated the basic musical grammar. The model of such communication can be found in the surviving tradition of Scandinavian pastoral music - kulning. This article discusses the most likely ways in which such music evolved.

**Keywords: tonal organization, animal communication, honest signal, semiosis, aspects of expression, domestication, kulning vs. yodel, motherese**

**Abbreviations:** AC, animal call (plural ACs); AE, aspect of expression (plural AEs); FF, fundamental frequency (plural FFs); ky, thousand years; kya, thousand years ago; SFTO, semiotically functional tonal organization; TO, tonal organization.

OPEN ACCESS

***Edited by:***
*Danielle Sulikowski,*
*Charles Sturt University, Australia*

***Reviewed by:***
*Colwyn Trevarthen,*
*University of Edinburgh,*
*United Kingdom*
*Mark Reybrouck,*
*KU Leuven, Belgium*

***Correspondence:**
*Aleksey Nikolsky*
*aleksey@braavo.org*

***Specialty section:***
*This article was submitted to Evolutionary Psychology, a section of the journal Frontiers in Psychology*

***Received:*** *21 November 2019*
***Accepted:*** *22 May 2020*
***Published:*** *23 July 2020*

***Citation:***
*Nikolsky A (2020) The Pastoral Origin of Semiotically Functional Tonal Organization of Music. Front. Psychol. 11:1358. doi: 10.3389/fpsyg.2020.01358*

## TONAL ORGANIZATION AND MUSICAL MODE

Since antiquity, scholars have been puzzled by the origins of music. Their quest still remains largely unanswered—impeded by the shortage of available data. The current consensus holds that some kind of musilanguage (Brown, 2000) must have preceded the bifurcation of music and language, marking the emergence of behavioral modernity in humans (Cross, 1999). Pitch orientation is seen as the primary structural marker of music, followed by rhythmo-metric organization (Brown, 2017)[1]. This unnecessarily oversimplified view can and should be expanded, since in reality music is organized not in two but in eleven *aspects of expression* (AEs[2]), each providing its autonomous information channel (**Table 1**):

- Melodic contour,
- Harmony,
- Texture,
- Form/thematicity,
- Tempo,
- Rhythm,
- Meter,
- Articulation,
- Dynamics,
- Register,
- Timbral quality (instrumentation)[3].

The problem is that in investigation of music, cognitive scientists rely on "standards" of Western musical theory, produced by Western civilization and therefore specific to certain historic periods and geographic regions. Although Western music system has proved to be the widest spread and the oldest surviving tradition, with its theoretic foundation rooted in the 3rd millennium BC (Dumbrill, 1998; Mathiesen, 1999; Jorgensen, 2003; Christensen, 2008; Crickmore, 2009; Nikolsky, 2016), nevertheless, there are other civilizations that abide by their own musical theories, explicit or/and implicit, documented or/and orally transmitted (Nettl, 2005). The need to formulate a "meta-theory" applicable to all varieties of musics has been realized only in the 1890s and dealt with by the discipline of systematic musicology (Bader, 2018). However, this discipline too inherited the framework of Western "classical music," which is just one of many (Nikolsky, 2015b, 2016, 2020; Nikolsky et al., 2020). Since this framework is tailored to incremental frequency changes, the pitch-related AEs have been prioritized in Western musicology, covered by the dedicated disciplines of harmony, counterpoint, and musical form (Christensen, 2008). The other AEs have only recently received attention, after the traditional discipline of musical form was approached semiotically (Bobrovsky, 1978; Mazel, 1979; Ratner, 1980; Nazaikinsky, 1982, 1988, 2013; Lerdahl and Jackendoff, 1985; Berry, 1987; Ruwet and Everist, 1987; Beliayev, 1990b; Molino, 1990; Nattiez, 1990; Aranovsky, 1991, 1998; Monelle, 1992, 2000, 2006; Narmour, 1992; Tarasti, 1994, 1995, 2012; Kholopova, 2002; Arom, 2004; Bonfeld, 2006; Medushevsky, 2010; Tagg, 2012; Turino, 2014; Benjamin et al., 2015; Yust, 2018). Cross-examination of syntactic, pragmatic, and semantic use of conventional musical idioms has revealed that they break into 11 different AEs (**Table 1**). Nine of them are used in monophonic music (without harmony and texture)[4]. Each AE is distinguished by its unique perceptual substrate and idiomatic expressions.

Interspecific comparison of human music to vocalizations of different animal species along these aspects promises a better understanding of the qualitative leap in the emergence of music. The Moscow school of "integrative analysis"[5] presents a methodology for such interspecific analyses, which I have adapted to identify those typological patterns in AEs of human music that contrast *animal calls* (ACs). These contrasts should be examined to reveal what exactly in human cultural evolution could be responsible for the emergence of new AE patterns that are unique to humans.

Human music is distinguished by its incremental structure (Bresin and Friberg, 2011)—requiring the ability to discriminate

[1]Metric organization of rhythm accompanies and supports pitch organization in music (Jones and Large, 1999), jointly supporting the "musical" manner of interpretation of sounds (Huron, 2006). Tonally important pitch-classes are usually stressed by longer durations and/or dynamic accents. However, comparing to rhythm, pitch organization is much more common in known world's music cultures (found even in music for only percussive instruments, e.g., African talking drums)—there are many forms of music that are characterized by ametric and arrhythmic free timing, but there are very few non-pitch forms. Therefore, pitch organization is a more reliable marker that distinguishes music from language than rhythmo-metric organization.

[2]I will use the abbreviation "AE" when speaking of a single aspect of expression, and "AEs" when speaking of multiple aspects of expression.

[3]The matter of choosing different timbres for different musical expressions has traditionally been handled by the discipline of instrumentation in Western classical music (Banshchikov, 1997). The term "instrumentation" here is somewhat misleading, because it covers not only the qualia of the timbres of musical instruments and their ensembles (trio, quartet, orchestra, orchestral group) but also various types of voices (soprano, tenor, bass), vocal ensembles (duet, trio, choir) and the rules of combining vocals with instruments (Kreitner et al., 2001). Arabic maqam, Persian dastgah, and Indian raga also observe similar rules in their respective practices.

[4]Technically speaking, monophonic music can still engage some idioms that relate to harmony and texture. A melody solo often features a pronounced "harmonic rhythm" (Swain, 2002)—i.e., periodic changes of implied chords (e.g., the "Blue Danube Waltz" theme by J. Strauss Jr.) that can stay regular (as in a metric pulse), be patterned (as in rhythm), or elaborated by expansion or contraction of a pulse period. Monophonic music can also implement changes in texture by patterning a stream of sounds into familiar textural idioms (e.g., the "Alberti figuration" or tremolo on a single tone) which then carry their specific semantic expression, different from other textural components, such as a melodic theme (Skrebkova-Filatova, 1985). However, overall, harmony and texture play a secondary role in monophonic compositional practices, limited to Western classical music alone.

[5]The "integralist school" of structural analysis of music was founded by the father of systematic musicology in Russia, Viktor Beliayev, in the 1920s, during his tenure in the Moscow Tchaikovsky Conservatory (Beliayev, 1990b). Beliayev's approach was further developed by two leading Moscow theorists, Leo Mazel and Viktor Tzukkerman (Mazel and Tzukkerman, 1967). They sought to integrate thorough structural analysis of a musical work with the psychological and sociological analyses of the expressive means employed in the analyzed musical work (Khannanov, 2005). It was especially Mazel who was concerned with broadening the framework of analysis to encompass not only the domains of melody and harmony, traditional for Western musicology, but also aspects of rhythm, meter, texture, articulation, dynamics, and timbre. After Yevgeny Nazaikinsky's death in 2006, the leading Russian "integralists" are Valentina Kholopova and Vyacheslav Medushevsky.

**TABLE 1 |** AEs of music.

| Acoustic domain | Aspects of expression in music | Physical substrate of an aspect | Perceptual substrate of an aspect | The overall range of expression of an aspect | Discrete increments within an aspect's range of expression | Gradual inflections within the continuum of an aspect |
|---|---|---|---|---|---|---|
| Frequency | 1. Melodic pitch (consecutive "linear" i.e., "horizontal") | Changes of the fundamental frequency (FF) between the consecutive tones within the same timbral register and voice/part. | Relation of the intervallic size of a frequency change to the Temporal Coherence Boundary[I] and the segregation of audio streams that result from it[LVI]. | Minimal change—about 100 cents, maximal change—about 1,200–1,600 cents (leaps over an octave characterize a few specific genres[XLVIII] (e.g., lamentation) or music systems (e.g., anhemitonic pentatonic). | The ambitus of a melody is divided into "degrees" based on the permanence in tuning (stability of pitch level) of tones of the same register that execute the same or similar melodic function within a phrase—the resulting intervals between the degrees define the "interval classes" in a musical mode[II]. | A degree can fluctuate in its frequency within a certain range of values that is usually equal or smaller than the interval between the adjacent degrees[III]—either in a form of a temporary alteration of that degree (syntactic inflection) or the portamento gliding between the adjacent tones (pragmatic inflection that traditionally constitutes the subject of intonation in Western music theory)[XLVII]. |
| | 2. Harmonic pitch (concurrent "vertical") | Consecutive changes in the relations of harmonics between the harmonic series of the concurrently sounding tones. | The number of harmonics that share the same frequency values in the harmonic series of each of the concurrently sounding tones[IV]. | Minimal matching of harmonics—when the distance between the FF of the concurrent tones is about or below 100 cents, maximal matching—when 1,200 cents apart. | The extent of matching and mismatching of harmonics varies between the degrees of a musical mode, forming a progressive "scale" of harmonic interval classes afforded by that mode, from the most "consonant" to the most "dissonant" interval[V]. | Each harmonic interval can slightly fluctuate in their exact tuning depending on the context of melodic and harmonic relations between the constituent tones within the same musical mode in a music piece, especially major and minor intervals—as a part of expressive tuning[XXX]. |
| | 3. Musical texture | A specific type of arrangement (vertical and horizontal) of all musical sounds within a musical work | The number and relations of familiar conventional structural components in grouping of tones, themselves forming stereotypes of arrangement specific to certain genres[XLIII]; varying along 3 axes: density, range[XLIV] and functions[XLV]. | The simplest case of texture is a monodic melody where grouping is restrained to shifts in melodic direction and leaps[XLV] without rhythmic contrasts (e.g., Paganini—*Moto perpetuo*); the most complex is the great number of numerous parts/voices that are diverse in their function (e.g., Debussy—*La Mer*)[XLVI]. | Each texture breaks into a number of "stream segments" at its surface level of perception[XLVII]; forming discrete components—"textural cells" used as bricks in constructing a texture by vertical (chords) and horizontal grouping (motifs) of various complexity, functionality and hierarchic relations—ascribed specific semantic values[XII]. | N/A |
| | 4. Musical form | Changes of the thematic material—a complex of musical structures consecutively ordered within a music work[XXXIX]. | Repetition, variation, contrast or recapitulation (i.e., the return of a thematic material after some other material) of a specific thematic material, identified by some salient feature(s)[XXXVIII]. | The simplest form is an exact reproduction of the same material (AA). The most complex is the "unveiling form" based on the ongoing contrast (ABCD. . .)[XXXIX]. | Changes in thematic material break a music work into discrete sections, with hierarchic relations between the changing phrases (A-B) and their changing constituent motifs (a-b), forming different hierarchic levels (e.g., A-B-A = ab-cd-ab). | Each section of a musical form can employ thematic materials of various salience, ranging from highly concentrated (symphony) to highly dispersed (prelude)[XLI]; the transition from the concentrated to the dispersed state can occur within the same section gradually[XL]. |
| Time | 5. Rhythm | Relative duration of consecutive tones is quantized according to a certain division ratio (2,3. . .). | Grouping of the consecutive tones based on their perceived proportions of duration and position within a group[VI]. | Minimum—according to the Western classical music theory, semihemidemisemiquaver, or, 128th note (=6th division of a whole note), maximum—brevis (=2 whole notes)—relative to tempo; in absolute terms, from 20 msec[VII] to 1,800 (2,000) ms[XXXI]. | Rhythmic proportions are estimated in terms of binary or ternary divisions that produce a set of standard durations—i.e., "time classes" (=rhythmic values) engaged in a composition—usually 3–5 divisions, one or two of which are most frequently used, forming a "metric grid" employed to round up the actual duration of a tone to a valid rhythmic value. | Rhythmic values can fluctuate in their actual duration in the so-called "expressive timing"[VII] that exaggerates rhythmic contrasts by prolonging anchored tones while shortening tones in passages, ornaments, or short tones in those rhythmic figures that consist of contrasting durations—e.g., overdotting in the so-called "punctured" rhythm[VIII]; such flections overtake the normative ratios that govern the rhythmic divisions[VII]. |

*(Continued)*

**TABLE 1 |** Continued

| Acoustic domain | Aspects of expression in music | Physical substrate of an aspect | Perceptual substrate of an aspect | The overall range of expression of an aspect | Discrete increments within an aspect's range of expression | Gradual inflections within the continuum of an aspect |
|---|---|---|---|---|---|---|
| | 6. Meter | Number of unstressed beats, grouped together with the stressed beat, before the occurrence of the next stressed beat—perceptually and statistically prevailing in a musical movement. | Grouping of beats based on the perceived periodicity of stresses generated by longer and louder tones as well as changes in melodic direction and harmony[VI]. | Minimal size group—1, i.e., "spondaic" pulse (every single beat is stressed), maximal size group—24 (compound ternary pulse, made by 4 divisions in 3-level hierarchy 3:6:12:24)[XXXII]; in absolute terms, periodicities from 100 to 6,000 ms[IX]. | Metric grouping is estimated in binary or ternary increments of an entire group that can be isochronous or non-isochronous[IX], so that a metric pulse within a meter can proceed by symmetric or asymmetric increments (e.g., common time can switch between 1/1, 2/2, 4/4, 8/8 and 16/16 pulses, or 3/8 + 3/8 + 2/8, 3/8 + 2/8 + 3/8 etc.)[L]. | Metric stress can fluctuate within a metric group, making the metric increments inside it acquire or lose metric weight: this is achieved by placing the anacrusis on a different position within a metric group (e.g., in 4/4 pulse, placing the anacrusis on the fourth beat shrinks the group to three beats—such shift alters the metric pulse without replacing it by another pulse[LV]. |
| | 7. Tempo | The average pace of beat within a span of musical movement, which retains a specific character of motion (e.g., hasty or lazy). | An overall impression of a characteristic movement of a certain type (e.g., walking, jogging, running, hopping) akin to a gait—estimated based on the interaction of the pace of beat and the rhythm which determines the choice of tempo by a performer[IX] enabling an "absolute tempo" (i.e., the optimal pace for a given music piece)[X] whose importance is reflected by the invention of the metronome[XI]. | Minimal metronome value—usually, 40 bpm, maximal metronome value—208 bpm, traditionally tied to heartbeat and gait rates that characterize a particular form of locomotion and its related affective state[XXII], in absolute terms, within the range of the beat value from 300 to 900 ms[VII]– inferred based on the density of pitch changes per metric unit of time[XXXIII]. | Each tempo is defined by a specific velocity and a character of musical movement (e.g., *presto* rushes, while *allegro* does not), so that every musical culture works out a set of standard tempi—for classical music of the Common practice period it is a 12-tempi system[XII] where a tempo, optimal for a given music piece, is defined as a range of bmp values within which the "feel" for that tempo remains the same[XIII]—narrow enough to consider a "perfect tempo."[XXXIV] | Velocity of each tempo can be adjusted without changing its character—music practice often generates rules for temporary minor fluctuations, reflected by a set of modifier terms: e.g., for Western classical music these are *meno mosso*, *ritenuto*, *rallentando*, *piu mosso*, *stretto*, *accelerando*; tempo inflections can also be canceled (*tempo giusto* = strict time) or added (*tempo rubato* = constant slowing and speeding within the same phrase)[XIII]; the velocity curves for such inflections seem to be fixed by convention in reference to the cultural standards of locomotion[XXXV]. |
| | 8. Articulation | The manner of attaching/detaching of successive tones within the same register and part/voice. | Shortening of the nominal rhythmic value (akin to the plucking sound production of mandolin or xylophone)—or, extending it (akin to echo) by the overlap of the end of one tone with the onset of the following tone (thereby generating a momentary harmonic interval in a monophonic line). | Minimal use of articulation—*non-legato*, maximal detaching—*staccatissimo*, maximal attaching—*legatissimo*[XIV]; these concepts do not seem to follow any absolute criteria. | Performance practice generates styles of consecutive rendition of textural elements (e.g., melody, chords, figurations) that fill a range from the most abbreviated to the most extended articulation in a gradient manner: *staccato*, *marcato*, *mezzo staccato*, *non-legato*, *portato*, *tenuto*, *legato*[XIV]—their contrast often generates groups (e.g., a 2-tone legato-tenuto or a 2-tone legato-staccato)[XLII]. | Many articulation styles form a continuous range of shortening or extending a rhythmic value, depending on the musical context—which establishes the performing conventions[XV]; a common case of contextual influence is adjusting the exact extent of legato[XVI] and staccato[XVII] depending on how high or low the register to which an articulated tone belongs is placed in the ambitus—the most common axes in flexing the articulation styles are connectedness, discreteness and compactness[XLII]. |
| Amplitude | 9. Dynamics | Changes of amplitude between consecutive or/and concurrent tones within the musical texture. | Relative increase or decrease in intensity of a particular tone, textural element (e.g., melody, bass), component (chord), segment (accompaniment) or the entire musical texture. | Minimal dynamics—*pianissisimo* (*ppp*), maximal—*fortissisimo* (*fff*); these concepts do not seem to follow any absolute criteria, yet are present in many if not all music cultures, from Ancient Greece on[XXXVII]. | Music practice generates dynamic distinctions that generally correspond to the extent of affective intensity of the music[XVIII] and form a "scale" of dynamics increments: *fortissimo*, *mezzo forte*, *forte*, *mezzo piano*, *piano*, *pianissimo*[XIX]—each featuring a range, narrower for experienced music users, highly variable in reproductions of the same music, but stabler per person[XX]. | Dynamics also uses gradual changes, usually to support climaxes and intensify contrasts[XXXVII]; such flections can be graded: positive—*piu forte*, *poco crescendo, molto crescendo*, *rinsforzando*; and negative—*meno forte*, *poco diminuendo* or *molto diminuendo*, *morendo*; although these terms appear in Western tradition only in the 19th century, similar notions seem to exist in implicit non-Western music theories to support flections of the expressive timing[XXI] that is most crucial for phrasing[XIV]. |

*(Continued)*

**TABLE 1 |** Continued

| Acoustic domain | Aspects of expression in music | Physical substrate of an aspect | Perceptual substrate of an aspect | The overall range of expression of an aspect | Discrete increments within an aspect's range of expression | Gradual inflections within the continuum of an aspect |
|---|---|---|---|---|---|---|
| Timbre | 10. Register | Contrasting changes in tonal quality between timbrally homogenous groups of tones within the ambitus of a music work. | Registral position in music is evaluated similarly to pitch—in terms of gradation in higher/lower placement within the ambitus that is employed in a music work—while accounting for timbral similarity in sound quality between adjacent pitches (e.g., registers that are darker/lighter or thicker/thinner in sound)[XXII]. | The lowest register forms one pole in the range of musical tones possible for vocal and instrumental production, while the highest register forms the opposite pole; for vocals and such instruments as flute or clarinet, the highest register is the strongest and most vibrant, whereas the lowest—the weakest and dullest; for brass, in contrary, the lowest register is the strongest and the most vibrant[LI]. | Human voices and musical instruments have break-up points in their tessitura, where the tonal quality noticeably changes, breaking in a few registers, each distinguished by its own coloration; thus, clarinet has four registers: somber "chalumeau" (E4-E5), dull "transition" (E5-B5), bright "clarino" (B5-C7), piercing "altissimo" (C7-A7)[XXIII]; 4 (3) registers are typical for most singing voices[XXIV] and musical instruments[XXV], but different cultures adopt different attitudes to registers: some smoothen registral contrasts, while others increase these contrasts[XXVI]. | Some musical traditions cultivate an overlap between the neighboring registers, extending the span of each register, thereby increasing its continuity—e.g., the countertenor can sing as baritone or bass[XXVII]; additionally, musical instruments usually develop an arsenal of performing devices to diversify their timbre providing "flections" of their "principal" timbre—e.g., *pizzicato*, *col legno*, *con sordino*, *sul ponticello* and *sul tasto* on string instruments[XXVIII]; similar devices are used by vocalists (parlando, aspirare, fioco); yet another common source of "flexing" the timbre and register is to stress a particular harmonic in the instrumental sound, thereby recoloring its timbre[XXIX]. |
| | 11. Instrumentation | Selection of a type of musical instrument and vocals most suitable for a specific expression | Timbres of individual instruments and vocals can blend into a new timbre (e.g., oboe and clarinet), remain discrete yet complement each other (e.g., flute and oboe), or repel (e.g., harp and horn), depending on similarity and synchrony of spectral centroids and attacks[XLIX]; and salience of individual harmonics[XXIX]. | The simplest instrumentation is sustaining a single timbre per piece of music (solo); the most complex is the combination of orchestra and choir that features multiple foreground and background layers, changing over time, with contrasts between tutti, soli, and orchestral and/or choral groups[LII]. | Each type of musical instrument and vocals constitutes a specific tone color in a palette of a music-maker[LI]; certain combinations of instruments (string trio, wind quintet, orchestra)[LIII] and vocals (duet, quartet, choir)[LIV] form stable settings used to create music of certain semantic content depending on the tonal quality and technical capacities of the instruments[LII]. | N/A |

*This table summarizes the structural characteristics of each of the principal AEs according to the treatises on music theory and relevant psychoacoustic research during the last 70 years. In the footnotes to the table, I have provided the relevant sources for those readers who are interested to find out more. All AEs feature incremental organization. Most AEs are quantifiable: melodic and harmonic intervals, textural elements and components, themes, metric pulses, rhythms, and articulation groups. Of 11 AEs, 9 (melody, harmony, form, rhythm, meter, articulation, dynamics and register) exhibit gradual inflections that are systemically used to increase the expressiveness of music.*

*[I](Noorden, 1975, 40–67); [II](Nikolsky, 2015b); [III](Garbuzov, 1948); [IV](Benson, 2007); [V](McDermott et al., 2010); [VI](Jones, 2016); [VII](Clarke, 2007); [VIII](Fabian and Schubert, 2008); [IX](London, 2004); [X](Levitin, 1994); [XI](Fallows, 2001); [XII](Nazaikinsky, 1972); [XIII](Garbuzov, 1950); [XIV](Keller, 1973); [XV](Jerkert, 2003); [XVI](Repp, 1995); [XVII](Repp, 1998); [XVIII](Dean et al., 2011); [XIX](Berndt and Hähnel, 2010); [XX](Garbuzov, 1955); [XXI](Todd, 1992); [XXII](Drabkin, 2001); [XXIII](Miller, 2014); [XXIV](Titze, 1988); [XXV](Patterson et al., 2010); [XXVI](Yemelyanov, 2000, 46); [XXVII](Ravens, 2014); [XXVIII](Garbuzov, 1956); [XXIX](Nazaikinsky and Rags, 1964); [XXX](Rags, 1980); [XXXI](Fraisse, 1982); [XXXII](Harding, 1983); [XXXIII](Madison and Paulin, 2010); [XXXIV](Gabrielsson, 1999); [XXXV](Friberg and Sundberg, 1999); [XXXVI](Chew, 2001); [XXXVII](Thiemel, 2001); [XXXVIII](Réti, 1951); [XXXIX](Mazel, 1979); [XL](Kholopov, 2006); [XLI](Val'kova, 1992); [XLII](Braudo, 1961); [XLIII](Huron, 1989); [XLIV](Benward and Saker, 2009); [XLV](Skrebkova-Filatova, 1985); [XLVI](Berry, 1987); [XLVII](Cambouropoulos, 2010); [XLVIII](Rags, 1999); [XLIX](Sandell, 1995); [L](Kholopova, 2002); [LI](Meyer, 2009); [LII](Banshchikov, 1997); [LIII](Kendall and Carterette, 1993); [LIV](Kreitner et al., 2001); [LV](Rothstein, 1989); [LVI](Cambouropoulos, 2008).*

changes in at least 9 AEs (**Table 1**). Their categorization into "classes" seems to be modeled after pitch. A music-maker breaks the range between the lowest and the highest pitch classes (i.e., ambitus) within a music work into "degrees," forming a set of pitch classes to construct music. Similarly, other AEs divide the continuum between their marginal values into step-like increments, the assortment of which can structurally characterize a musical work. Pitch-class sets receive their analogs in sets of the following classes, intuitively selected by a music-maker for a particular expression per composition:

- "time-classes" (number of rhythmic values i.e., "divisions"),
- "pulse-classes" (number of periodicities in a metric grid),
- "tempo-classes" (number of musical movements)[6],
- "articulation-classes" (number of styles of connecting consecutive tones),
- "dynamics-classes" (number of dynamic gradations),
- "register-classes" (number of zones of different tonal coloration),
- "texture-classes" (number of textural components),
- "form-classes" (number of themes).

Such discrete classes coexist with gradual inflections for each class (**Table 1**). Evidently, music is designed to integrate multiple AEs in a complex admixture of their patterns of expression. Music defaults to the *integration* of concurrent tones in contrast to the *segmentation* tendency of speech (Bregman, 1994)—people can sing together, yet when speaking, they always take turns (Brown, 2007). Here, AC sides with music rather than speech, evident in the widespread animal chorusing. Integrative power of music makes the concept of "musical mode" indispensable for understanding the rise of music. "Mode's" reduction to "scale," adopted by some researchers (i.e., Pfordresher and Brown, 2017) constitutes a fundamental error in confusing the purely quantitative and formalistic concept of "scale" with the qualitative and content-oriented concept of "mode" (see Nikolsky, 2015b). Musical mode is more than a mere set of pitch-classes selected to make music—it also encapsulates the rules for their interconnection and the semantic range of suitable expressions (Wulstan, 1971; Alekseyev, 1976; Kholopov, 1976, 2005; Bytchkov, 1987, 1997; Lester, 1989; Beliayev, 1990a; Porter et al., 2001; Powers and Wiering, 2001; Straehley and Loebach, 2014; Winnington-Ingram, 2015).

In essence, "mode" constitutes the generalization of a particular melodic typology, characteristic for a given musical genre, which supplies that mode with semantic denotations (Nazaikinsky, 2013). Nothing similar exists in speech. Music is unique in its holistic appreciation of sounds *per se* (Patel, 2010). Hence, the idea of *euphony*—pleasant concordance of sounds in specific expressions—is quintessential for "mode," as emphasized by Russian theorists.

The same principles apply to "rhythmic modes," conceptualized within Western (Roesner, 2001) and some non-Western civilizations (Clayton, 2000). Rhythmic divisions, utilized in a composition, complement one another in expression of musical movement and in combinatory rules. A rhythmic modus in Western medieval theory, Arabic maqam, Iranian dastgah, or Indian raga incorporates not only a specific progression of rhythmic values but a specific "ethos"— an abstracted emotional quality projected by music on society at large (Shestakov, 1975). Each rhythmic modus in the abovementioned music systems is characterized *semantically* by its affiliation with a certain ethos and *structurally* by certain proportions between the duration values used in a music work. Rhythmic modus resembles pitch modus by incorporating a set of rules. Just as pitch-classes are allowed to follow or not follow one another, or require an alteration for ascending or descending motion, rhythm-classes are restricted to certain ratios which can be altered in a certain way (e.g., a dotted rhythm can be "over-dotted" in a suitable context).

The idea of concordance and appreciation that underlies the overwhelming majority of known traditional music cultures justifies the conceptualization of each AE as a carrier of its proprietary "mode." Every musical piece can be defined by identifying its melodic, harmonic, rhythmic, metric, tempo, articulation, textural, and timbral modes.

Together, these modes constitute "tonal organization" (TO) in music. Conceptualized by François-Joseph Fétis (1840), TO is a method of joining musical tones together according to the sensibility of music-users (Fétis, 1994, XXV). Unlike tonemes of tonal languages, musical TO affects *all* tones, generates *complex* functional relations between them, and involves rhythmo-metric, dynamic, articulatory, and registral arrangements. Speech might also use similar arrangements (Patel, 2006). But music requires a special analytic attention where changes in the melodic contour are quantized into pitch-classes that are continuously cross-compared—unlike the linguistic "vowel pitch" (Walker, 1997, 322–3). Such syntactic pitch-parsing is as imperative for music as word-parsing is for language. Semantics provides yet another distinction: verbal syntax specializes in conveying referential meaning, whereas music specializes in emotional expression[7] (Gabrielsson and Lindström, 2001; Juslin, 2001, 2005, 2011, 2013; Cook, 2002; Krumhansl, 2002; Gabrielsson and Juslin, 2003; Dissanayake, 2008; Johnson-Laird and Oatley, 2010; Trainor, 2010; Perlovsky, 2012; Altenmüller et al., 2013b; Eerola and Vuoskoski, 2013; Eerola et al., 2013; Peretz, 2013; Nikolsky, 2015a, 2020; Schiavio et al., 2016). Such distinction has been

[6]The word "movement" here refers to a principal division of a longer music work into sizeable sections, each distinguished by its own metric organization and tempo: e.g., a 4-movement symphony or a 3-movement sonata. The concept of movement emerged in 16th-century Western classical music to reflect on the old practice of switching from one tempo to another within the same piece of music (Sadie, 2001). However, by no means the use of multiple movements within the same work is exclusive to Western civilization. Well known are non-Western genres of music that employ cyclic arrangement, such as Arabo-Andalusian nubah (Touma, 1996) or Javan court Gamelan music (Sutton, 1991).

[7]Ontologically, it is necessary to distinguish between "meaning" in a natural language and "meaning" in a cultural system of symbols (such as music)—especially in light of the difference in their acquisition: thus, under experimental conditions non-human primates can acquire some symbolic systems but not a full-fledged human language (Balari et al., 2011). It seems that the verbal combinatorial semiosis of referential meaning is fundamentally different from retrieving imagery, be it emotional or motivational information assigned to cultural symbols. This distinction is crucial for the investigation of origins of human language and music. Here, music, despite its combinatorial nature, occupies a place closer to signal-like semiotic systems, which *makes music more accessible to hominins than language*.

fundamental for the musical practices and theories of most musical traditions before Western classical music was swept away by the 20th century modernistic "revolution." This distinction became revived after emotion and music attracted intense neuro-psychological research in the 1980s.

Music's social nature—evident in entrainment[8] (Tarr et al., 2014)—and emotionality—evident in chills (Altenmüller et al., 2013a)—are critical for distinguishing music: neither entrainment nor chills characterize verbal communication. And both are closely related through emotional contagion (Trost et al., 2017). This music/language distinction must have been already present in musilanguage, since in AC referential and motivational information is coded differently (Manser, 2010). However, music differs from ACs by encoding affective information according to the conventional modes of numerous AEs, as we shall see. Hence, the structural definition of music should be:

TO of multiple AEs that *entrains* listeners and performers and *transposes* performers' intentions to emotionally stir listeners through vocal and/or instrumental performance.

Pitch contour, rhythm/meter, and dynamics (the most salient AEs) together constitute the principal structural criteria of music.

## EMIC AND ETIC APPROACHES TO TONAL ORGANIZATION

The proposed definition is instrumental for engaging an additional source of evidence in the quest for the origins of music—the comparative structural analysis of world's archaic indigenous musics, earliest forms of music-making by human infants, and animal vocalizations. The modern advances in computer science support the acoustic and statistical analyses of vast datasets unavailable before. Such investigation could radically update the evolutionary theory while resolving the current situation in comparative ethnomusicology that is nothing short of a crisis (Savage and Brown, 2013).

Many cognitive scientists remain unaware of the profound ideological shift in Western ethnomusicology that occurred during the last half-century. In essence, the study of "text" became replaced by the study of "people" (Zemtsovsky, 1997)[9]. The turning point was marked by Gourlay (1982) at the 1979 Oslo Conference of the IFMC by a call for "humanizing ethnomusicology" to abandon "the pretense of objectivity." Timothy Rice reflected this departure in his influential article "Remodeling Ethnomusicology" (Rice, 1987). At the heart of this transformation lies the emic/etic antithesis, introduced by Pike (1967) in 1957 to oppose the "insider's" versus the "outsider's view" in the researcher's position toward an object of study. Ever since, this opposition has grown into a schism between Western social and cognitive scientists (Headland, 1990). Harris (1964) adapted Pike's approach for social sciences, conceptualizing "emic" as a specific culture, *mentally* "native" to an "insider," whereas "etic"—as cultures, experienced not mentally, but *behaviorally* due to their "foreignness" to an "outsider." Hence, Harris' claim that an outsider is capable of only grasping the superficial behavioral patterns through direct observation. Harris' followers wanted to *abstain* from any "mentalization" of observed facts to avoid their misrepresentation (Harris, 1990). Pike's followers, in contrary, interconnected mental and behavioral aspects, holding that etics and emics present respectively physical and cultural aspects of analysis, so that an outsider *can* learn to analyze like an insider, and vice versa (Pike, 1990).

For ethnomusicology, emic/etic problem was discussed at the 32nd ICTM Conference, 1993, Berlin. The consensus recognized that insider and outsider perspectives were inseparable and complementary to each other: emic data was to be fit into etic categories, disregarding whether they were actually recognized by the insiders (Baumann, 1993). However, in the following decade Western ethnomusicology became progressively politicized against a supposed "Western bias"—equated with any form of etic evaluation. Some authorities went as far as viewing cross-cultural scientific investigation of music as "cultural colonialism" (see Agawu, 2003).

The purist emic approach replaces the scientific method of investigation with the insider's description of a native culture in a social context (Myers, 1993, 222–3). The reason for this is that the scientific method by itself is a product of Western civilization (Messner, 1993). Thus, Gourlay (1984) explicitly defies any objective inquiry about music by means of scientific investigation[10]. Becker (1986) declares musical systems as being "incommensurable," and any scientific study of non-Western music as being "immoral." She insists that each musical culture should be investigated only in its own native terms and *not* evaluated against another culture—the only way for a researcher to study music is to merge with the indigenous community, learn its language and jargon, and collectively make music. In effect, this utilitarian ethno-unilateral approach to music precludes the study of its origins (Dobzhanskaya, 2012). No wonder, in the West, comparative musicology became abandoned,

[8]Entrainment (from French "*en-*" + "*traîner*"—to drag something along) is the term used in physics to address a wide range of phenomena where two oscillators are coupled, and one of them gradually comes into synchrony with the other, becoming locked in a phase. Entrainment of two pendulum clocks was discovered by Christiaan Huygens in 1666 but was explained only few centuries later. In early 20th century, other manifestations of entrainment were unveiled in acoustics (coupling of the organ pipes) and biology (glimmering fire-flies)—until it was generalized as a universal physical phenomenon (Pikovsky et al., 2001). Its biomusicological manifestations were identified in the 1990s, at first in relation to music therapy, and thereafter as an integral part of perception of rhythm and meter (Large and Kolen, 1994), of great importance to the evolution of music (Fitch, 2012).

[9]Thus, Titon (2015), one of the leading Western ethnomusicologists of today, goes as far as defining the discipline of ethnomusicology as "*the study of people making music*"—rather than "the study of music" as the term "musicology" indicates (the study of human societies is conducted by another discipline—"anthropology," reflected in the etymology of its name). Paradoxically, modern Western "people's ethnomusicology" still shuns the Soviet ethnomusicology which shared the same approach, holding music as "belonging" to people and "reflecting" people's mentality, while remaining totally free of the anti-textual bias (Panteleeva, 2019).

[10]Gourlay argues that *no musicological study* of African music *by outsiders is justified*, because "in no African language about which we have information, and in many used by other peoples who have oral rather than written traditions, is there a word corresponding to the English term 'music'." So, according to Gourlay, "where the term 'music' is unknown to the people in question, one can conclude only that what we are presented with is *the investigating scholar's* concept of *his/her* 'music'."

musical universals denied, and music history fragmented into a bunch of disconnected "histories" (Savage and Brown, 2013). Unfortunately, despite its severe shortcomings, the "emic bias" has penetrated into psychoacoustics (i.e., see Parncutt and Hair, 2011)[11].

Certainly, not all Western ethnomusicologists abstain from the musicological analysis (Arom, 2010) and deny the validity of objective etic approach (Alvarez-Pereyre and Arom, 1993). Nevertheless, the anti-analytical trend[12] has taken its toll, establishing a conviction that any research of structural universals is inevitably ethnocentric and inadmissible for ethnomusicology (Nattiez, 2012). Disregarding musical text in sake of musical behavior is symptomatic of a shift away from comparative musicology to fractured sociomusicology of isolated musical communities (Nettl, 2010, 70–92). Many contemporary American ethnomusicological papers are published without a single example of structural analysis to support the author's claims, basing their claims on entirely behavioral, and not musicological, data—paradoxically conducting *musicological* research without looking into *music per se* (Zemtsovsky, 2002)[13]. Consequently, cognitive scientists interested in comparative music theory and musicological analysis have no choice but to rely on the old publications in English and new ones in other languages (especially those coming from Eastern Europe and Asia, where the influence of politicization is weaker).

The summary of etic/emic arguments, crucial for investigation of TO, demonstrates that proponents of emic approach strongly overvalue it while writing off its fundamental flaws (**Table 2**).

TO is identifiable based on the etic information alone, and its few potential shortcomings are easily amendable by emic references (Dasen, 2012). Purely etic approach has been a status quo in organology, where musical instruments are identified according to etic principles, disregarding emic views (Baumann, 1993). And there is no reason why the entire field of ethnomusicology should not be treated in the same way. The etic approach is unique in enabling a "progressive" accumulation of knowledge where the mistake of one researcher can be corrected by another. Etic self-sufficiency is evident in the fields of ethology and developmental psychology. Neither human babies nor animals can provide emic information—which by no means invalidates the acoustic analysis of their communication.

In light of this, studying TO is paramount for establishing the objective ground for interdisciplinary scientific research of the evolution of music across the synchronic and diachronic varieties of music systems. TO's role for musicology is comparable to the role of phonology in linguistics: TO specifies a set of acoustic attributes and their oppositions to encode and convey information. Together, they form the "surface level" that underlies the musical syntax and semantics, and provide the material base for any music culture (Cambouropoulos, 2010).

## TONAL ORGANIZATION DISTINGUISHES HUMAN MUSIC FROM ANIMAL COMMUNICATION

The very ability to enjoy "harmonious" sounds most likely emerged as a byproduct of satisfying the need to bring individual emotions in accordance with the interests of a social group (Panksepp and Bernatzky, 2002). Musical anhedonia in humans is exceedingly rare, indicating that music evolved as a direct auditory pathway toward the emotional reward centers in the brain (Loui et al., 2017). Music is probably a human invention that came-into-being to shape important brain functions through the hedonistic effect of appreciating sounds (Patel, 2010). Patel's (2008) theory of "transformative technology of the mind" reconciled the adaptionist (Darwinian) and the non-adaptionist (Spencerian) approaches, based on the latest cognitive research, and provided the foundation for the theory of "mixed origins of music" (Altenmüller et al., 2013b) that explains how human affective signaling system has transformed the human brain and created music. Emotive specialization and emergence of "musical emotions" must have followed the formation of human auditory-affective circuitry (Altenmüller et al., 2013a).

Centrality of affective signaling brings animal communication closer to music than to speech (Fitch, 2006). Animal signals usually express affective states according to their innate "vocabulary," are volitionally produced, and are actually felt (Fitch, 2010, 179–81). TO shares more similarities with animal vocalizations than with phonetics, since consonants, crucial for verbal parsing, are unique to human speech—unlike vowels that are more similar to singing and ACs (Kolinsky et al., 2009). Vowels determine verbal prosody which is the primary means of conveying emotions through speech.

Most likely, the musilanguage's TO resembled the model of vocal production, common for primates and human infants—a reflex-like vocalization (e.g., pain-shrieking), triggered by specific stimuli, and hard-wired for animals but modifiable for humans (Jürgens, 1995). Humans start developing the repertory of cries by differentiating timbral and contour features just a few months after birth (Wermke and Mende, 2009), whereas for most animals, call structure is not modifiable by acoustic experience

[11] Parncutt and Hair subscribe to Gourlay's defiance of a scientific investigation for those phenomena that do not find a corresponding term in a native language. They categorically insist that the *research of consonance and dissonance be constrained only to music of such cultures that define the concepts of consonance and dissonance*: "if musicians in that culture do not talk directly or indirectly about C/D [consonance/dissonance], it is considered irrelevant." By this logic, there is no gravity in those countries whose native people do not have a word translatable in English as "gravity." Parncutt and Hair see the goal of studying music in "documenting the musical and music-theoretical discourses of the insiders about which tones and rhythms should be played together and why, and considering the political and psychological mechanisms that are allowing Western music to dominate world music"—undoubtedly, a controversial and a politically biased agenda.

[12] To substantiate this criticism that is rarely voiced in modern Western literature, I shall quote one of the biggest authorities in ethnomusicology (the emphasis is added by me): "*Functional analyses of musical structure* cannot be detached from *structural analyses of its social function*: the function of *tones in relation to each other cannot be explained adequately as part of a closed system* without reference to the structures of the *sociocultural system* of which the musical system is a part, and to the biological system to which all music makers belong" (Blacking, 1974, 30–31).

[13] One of the main reasons for the drop in standards of musicological and ethnomusicological analyses is that in the US and UK academic curricula, music theory in general, and music analysis in particular, have been offered as rudimentary undergraduate courses (Agawu, 2004). In contrast, in countries of the former Soviet Union, music analysis has been taught at the highest level of scholarship that requires at least 10 years of study before attaining a level of training where an analyst is expected to capture and interpret the totality of expressive means employed in a music work (Khannanov, 2005).

**TABLE 2 |** Pros and cons (P/C) of purely etic, emic, and combined "etic + emic" approaches to analyzing music structures.

| Task | Emic P/C | Etic P/C | Etic + Emic P/C |
|---|---|---|---|
| Categorization | Minimal distortion of the original concepts underlying music-making and adequacy of applying theory to practice | Imposing culturally foreign theoretical concepts onto indigenous music practice, and misunderstanding of its typology | Adequate categorization |
| | Possibility that an emic expert would not recognize an important category that is objectively evident in a given culture - due to its novelty and its contradiction to the traditional ideology or a dominant practice | Accurate categorization of all audible musical features—even if for some religious, ideological, historic, or ethnic reasons they are not noticed, ignored, misinterpreted, or forgotten in an indigenous music theory | Emic categorization should be discarded |
| | Possibility that an entire culture (e.g., herding vocalizations) or certain genres (spells) that appear as music to outsiders might be viewed as non-music by locals for religious, ideological, or practical reasons | All forms of musicking that appear as "music" to an outsider and satisfy the criteria of music are recognized as such, even if local experts for some reason do not consider them as music – thereby making the scope of study wider | Emic categorization should be discarded |
| Capturing important data | Correct identification of all the relevant data as long as it is adequately reflected in emic music theory and recognized during audition by an emic expert | Possibility of omitting information that is crucial for a local music culture, but totally unfamiliar to an outsider | Adequate data collection |
| | Omission of features of TO that have not been perceived by an emic expert (e.g., due to hearing or performing problems), conceptualized by the local emic theory, rejected as foreign or inappropriate, or misinterpreted in terms of the ideologically dominant tradition | Complete and comprehensive collection of all the data that is physically present in local music practice and whose typology has been known in other related or similar musical cultures | Emic manner of data collection should be discarded |
| | Omission of those genres or styles of music that are tabooed within a culture (e.g., funeral, totemic, or ancestral), which can severely distort the perspective on such musical culture | Acceptance of observable genres or styles tabooed for insiders by the outsider researchers, allowing them to integrate this information with the rest of what is known about this culture | Emic taboos should be discarded |
| Examination of data | A local expert has no problem understanding and interpreting the familiar musical material | Either an outsider must acquire the basic knowledge of local music material, or an insider has to learn some general theory | Adequate analysis |
| | Possibility that local music theory can be culturally or ethnically biased or tainted in relation to specific kind of data, and local experts would rigidly discard this data from any consideration in their analysis | Uniformity of the analytic approach to the musical material that characterizes a given culture secures equal treatment of all the collected data, making the findings more comparable with other findings | Emic methodology should be discarded |
| | The data pool for examination available to any single local expert is severely limited by their personal experience, and therefore, is prone to lead to erroneous generalizations | The use of standardized procedures enables automated data analysis, greatly expanding the data pool, and increasing the reliability of the findings | Emic data can be used only if it does not contain inaccurate generalizations |
| | An insider expert is prone to errors due to accidental lapses in attention – and these errors are hard to detect and correct | Standardized procedures secure the use of computer technology and collective work of numerous researchers, preventing and correcting accidental errors | It will be safer to discard the emic data unless it is confirmed by another expert |
| Verification of results | Results of the emic expert's analysis are verifiable by interviewing the insiders and conducting the statistical analysis of their responses, which can identify a problem but cannot resolve it | Results of the etic analysis cannot be directly verified by the indigenous music users, requiring a qualified interpreter and limiting the reliability of statistical analysis of the obtained data | Adequate verification |
| | Insiders might follow a system of irrational beliefs (i.e., spiritual) that preclude any rational verification of the results of emic-based TO analysis | Irrational criteria can be separated from the rational, and both can be effectively verified: rational by testing, and irrational by interviewing | Emic evaluation should be discarded |
| | There is no reliable way to reconcile any disagreement between two disagreeing insider experts | Disagreements between informants are rationally resolvable by testing more subjects and by conducting additional interviews, as needed | Emic evaluation should be discarded |
| Comparative Study | Natural coherence of a local culture as regards to ideological, structural, social, and semantic data facilitates its cross-cultural comparison, if the compared cultures are similar in composition | Reliable evaluation of a local culture is possible only in those domains that are physically manifested, and match the researcher's experience – otherwise, a comparison might be erroneous | Adequate comparison |
| | As a rule, cross-cultural comparisons are impossible due to the general mismatch of categories and definitions that vary between different emic music theories, both verbally and conceptually - even between linguistically related cultures: there is no way for emic experts in each of the cultures to secure a common ground | Cross-cultural comparisons are easy and comprehensive if the general music theory is equipped with means to identify and classify all the principal varieties in acoustic attribtues of tonal organization that are encountered in each of the local music cultures; then, the generalized category will secure a common ground for cross-cultural comparison | Emic comparison should be discarded |
| | To compare a few cultures, a researcher has to learn the conventions of music making, language, and customs of each culture, which is extremely time consuming and still does not guarantee adequate understanding of each culture | To compare numerous cultures, an outsider is not obliged to visit and stay in a given location, learn to make music in emic way, and speak the native language, which spares time and enables learning about a greater number of cultures | Emic comparison can be discarded |

*Pros are colored blue, and cons red. The number of cons for the emic approach (11) doubles the number of cons for the etic approach (5). Emic cons are more detrimental for the outcome of the analysis. The etic approach, even at its worse, still allows the researcher to infer valid principles of TO in a sufficiently large pool of samples of musical styles/genres - which, in the long run, secures correction of mistakes by subsequent researchers. At its worse, the emic approach precludes any comparative study and invalidates the study of TO in an isolated music culture where its members do not regard certain sound production as "music" (i.e., incantations, spells, herding vocalizations). The combined etic/emic approach effectively corrects the shortcomings of a purely etic approach, but in most cases, it fails to correct the shortcomings of a purely emic approach.*

(Hauser, 1996, 315). Call-learning occurs in a few songbird species, but for most birds, songs are innately encoded, and life experience only activates their retrieval (Marler, 1997).

A call serves as the basic unit in animal communication[14] and usually conveys specific affective information (Hauser, 2000). Different calls are combinable in "mixed bouts" that are different from "pure bouts" (single call) by triggering a sequence of emotion-based behavioral responses in other animals. Each call's significance is hard-bound to its acoustic structure. Despite their superficial similarity with music, "mixed bouts" lack transposability of intentions: each call comes only in response to the actual stimulus present in the environment (Zuberbühler, 2017). Transposability is the landmark of music—the same structural pattern is intended to express the same idea across different instances of use, without which musical genres would be impossible: e.g., most lullabies are recognized cross-culturally by their set of structural features (Trehub et al., 1993). Genres are based on reproduction and transposability, and usually form genre systems to support important social practices (Samson, 2001), which enables music to reflect perceptual reality. Animal-learned vocalizations miss such comprehensiveness and generalization. They are limited to:

- display of fitness (Naguib and Riebel, 2014),
- a single season and gender (Slater, 2011),
- mating or defending situations (Slater, 2001).

Syntactically, AC overall lacks a combinatorial organization[15]. It resembles the one-word holophrasic communication of human infants by depending on a directly observable context and on an "analog" signal-emotion correspondence (Johansson, 2005). The same applies to animal "*phonocoding*"[16] (Marler, 2001): it excludes categorical perception, rhythm, hierarchical structure, and adjacent transitional probabilities (Yip, 2006).

Indispensable for speech and music, compositionality completely eludes ACs—along with listener's capacity to continually (re)-organize behavior as the song unveils. Non-human communication, as a rule, employs a "one-ended" system: a signaling animal emits a signal unconsciously, not for any specific receiver but as a physiological reflex conditioned to a particular type of stimuli (Hauser, 2000). Such intention-free transmission precludes semiosis[17]—since sender and receiver must share signs and codes to actually transmit information.

A cumulative "two-ended" semiosis, where the receiver signals in response to the sender and vice versa, is unique to humans, and emerges as a result of technological complexity of human life. Dennett (1983) called this "second-order intentionality"—i.e., the receiver's beliefs and desires about the sender's beliefs and desires—in distinction from the "first-order intentionality" that is limited to the receiver alone.

- *First-order* intentionality is characterized by a one-ended conscious processing of unconsciously emitted signal—here, the unintended signaling receives an intentional interpretation.
- *Second-order intentionality* requires a two-ended premeditation of a signal: the signaler has to consider the receiver's competence, and the receiver must be looking for information while considering the signaler's circumstances.

Subsequently, the state of knowledge is changed on both ends of such communication, which, so far, has not been found in any non-human animal. Most common for ACs is *zero-order intentionality*—the signaler does not consciously intend to convey a piece of information, but *instinctively* engages a specific signal structure, triggering a similarly automatic response of the receiver.

Two-ended communication generates an unlimited diversity of structure due to infinite recombinations of a finite set of discrete elements that do not carry meaning on their own—what Abler (1989) calls "particulate principle." It is peculiar to human language and music, finding only embryonal equivalents in a few animal species (Hauser, 2000). Complexity, comparable to human, is evident in some birdsongs, but serves to impress mates and intimidate competitors rather than conveying a specific message (Marler and Slabbekoorn, 2004)—likely forming a parallel (not prototype) to human evolution (Fitch, 2010, 184).

[14]In some songbirds, the innate encoding consists of smaller elements, resembling syllables, and following simple rules for how to order them, so that a bird actually learns to "assemble" its song. However, the assortment of such elements is very limited, making songs signal-like, restrained to a single species. Playback of isolated syllables of such songs either does not elicit response or produces a weak reaction in other conspecific birds (Searcy, 1992). Perhaps the rearrangement of elements constitutes not a pragmatic, but a "syntactic" production unit—thus, zebra finches were found to stop at syllabic breaks in a song, when detracted (Cynx, 1990). Rearrangement of "syllables" is also used by a few primate species (gibbon) to disclose the identity of a caller for conspecific animals (Marler and Mitani, 2008).

[15]Although it is not uncommon for ACs to form a sequence according to a rule-based structure, noticeable by conspecific animals (Fitch, 2010, 182), changes in such structures apparently do not result in the changes of meaning of the entire song (Hauser, 2000). The most syntactically elaborated bird and whale songs use combinatorial features, albeit minimal. However, despite having a componential structure, such animal song in its entirety presents a *single* piece of information learned from the animal's parent holistically rather than incrementally, element by element (in contrast to how humans learn), and is therefore highly stereotypical in form (Hurford, 2012, 3–99).

[16]The concept of phonocoding (i.e., "phonological coding") was introduced to oppose "lexicoding" of human speech (Marler, 2001). Phonocoding refers to the capacity to generate new sound patterns by recombining the constituent elements and components of known conventional signals. This capacity is minimal in non-human primates, but common in learned vocalizations of songbirds and whales, which, however, remain primarily non-symbolic and affective.

[17]The term "semiosis" here refers to the Peircean concept of conveying information by encoding it into signs by one party and decoding it by another party—a "two-ended" system. A "one-ended" call can be somehow interpreted in relation to the situational context by the listening animal, but this interpretation can radically differ from the actual state of the sender: e.g., bird's mating call might be interpreted by a nearby cat not as a signal of readiness for mating but as a signal for hunting. Then, the integrity of the information passed from sender to receiver is not preserved. Within this context, the use of the term "meaning" in regard to an AC, adopted in biosemiotics (Sebeok, 1994, 111), is confusing, since "meaning" implies that someone "means" something by displaying a specific sign. More accurate here would be to employ the term "*significance*" (as in "to signify") instead of "meaning."

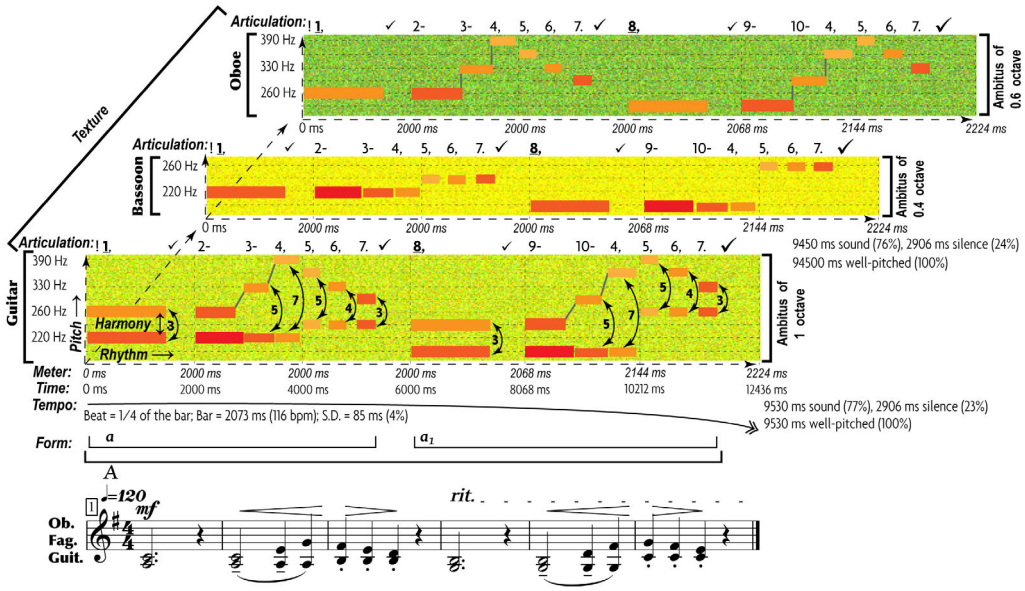



**FIGURE 1 |** 11 AEs in a musogram of classical instrumental music. At the bottom of the figure, the conventional musical notation represents the same content as the three musograms above it. The lowest musogram (guitar) contains all the AEs marked out and named. Its horizontal axis (horizontal dashed arrow) represents time, vertical axis (vertical dashed arrow) frequency, depth axis (diagonal dashed arrow) the aspect of **texture**. The latter joins all three musograms. Small colored rectangular bars indicate tones. Their vertical relation represents **pitch**, with dash guidelines referencing frequency values. The changes in distance between the concurrent (superimposed) rectangles indicate **harmony**. The rectangular length represents **rhythm**. The breaks and the gray lines that connect the consecutive rectangles as well as the numbers above the frequency grid comprise an aspect of **articulation**. Each tone is numbered, checkmarks indicate pauses (the bigger the pause the larger the checkmark), and punctuation signs reflect the grouping of tones. Dashes mark the connected tones (legato), commas—disconnected tones within the same phrase, periods—the end of a phrase, and exclamation marks—the phrasal opening. Bold and underlined numbers indicate anchor-tones (stressed by duration, dynamics, and frequency of occurrence). The gray lines represent connectivity: discrete pitches are connected by vertical lines, whereas portamento pitches by tilted lines. The coloring of rectangles represents **dynamics:** from the loudest in yellow to the softest in blue. Thin vertical dashed lines indicate **meter**—inferred from well-articulated occurrences of anchor-tones and longer rests. **Tempo** averages all metric units, expressed in msec and beats-per-minute. The standard deviation shows how flexible the tempo is. A solid arrow with a double arrowhead reflects the tempo changes: ascending for accelerations, while descending for decelerations. **Form** reflects the thematic organization of the material, indicated by horizontal brackets and letters: thinner brackets and lowercase letters for motifs, and thicker brackets and uppercase letters for phrases. Each new material is marked by a new letter, and variation—by a subscript number. **Register** is represented by the coloration of the grainy filling of the ambitus: from a deeper green for the darkest timbre to yellow for the lightest timbre. In this example, oboe uses its darkest register, bassoon—its faintest register, whereas guitar—its medium register. **Harmonicity** (see **Table 3**) is indicated by the relative thickness and the geometric shape in representation of tones: the greater the harmonic richness, the thicker the rectangular bars, whereas the noisier the sound, the more irregular the fuzzy shapes (not present in this particular example). For thorough explanation of this method of visualization see Appendix 1 in **Supplementary Material**.

The structural criterion for emergence of the *Semiotically Functional TO* (SFTO)[18] in music is therefore manifested in the introduction of particulate organization in *phonocoding*.

## THE TIMEFRAME OF TONAL ORGANIZATION OBTAINING FULL SEMIOTICALLY FUNCTIONAL CAPACITY

The current consensus holds that music was gradually formed since the appearance of Homo heidelbergensis about 600,000 BP, leading to an artistic "explosion" circa 40,000, when the earliest bone "flutes"[19] were produced "en masse" (Morley, 2013, 219–25). Although flutes prove the existence of TO in the Aurignacian culture, this tells nothing of whether their sounds served a one- or two-ended communication. In all likelihood, TO did

[18]By "semiotically functional," I mean that a music-maker selects the elements and components of tonal organization for each of the aspects of expression in music based on their efficacy in conveying specific affective information ("musical emotion") to his/her listeners and/or partners in performance. In this sense, the AC can be considered "semiotically dysfunctional"—not supporting a successful two-ended communication (delivery of the intended message) between the sender and the receiver.

[19]The word "flute" here is used informally: there is not enough archeological evidence to conclude if the earliest instruments were flutes or clarinets. The oldest artifact is a bone fragment from Haua Fteah, Libya, with a single hole, dated 90–110,000 years ago (Blench, 2013). Most archeologists do not recognize it as man-made. Next in line is the 47,000 years old 3-hole artifact from Divje Babe, Slovenia, uncovered in 1995. It was interpreted as a bone bitten by a carnivore (D'Errico et al., 1998). However, experimental testing has demonstrated that none of the cave bear, wolf or hyena dentition could punch two holes without cracking and splitting the bone (Turk et al., 2001). Nevertheless this argument was not accepted by the supporters of non-human origin of the Divje Babe artifact (Morley, 2006). Subsequent tomographic analysis has concluded that the Divje Babe artifact was man-made (Tuniz et al., 2012). Slovenian researchers have presented additional reasons for its man-made origin (Turk, 2014). In spite of this, another recent British study has restated the bite origin hypothesis (Diedrich, 2015)—though, without addressing the 2012 and 2014 studies' arguments. The third in timeline and unequivocal in its provenance, is the 5-hole Hohle Fels-1 flute, 35,000 years old (Conard et al., 2009).

not *communicate* musical emotions but merely *accompanied* the behavioral display of actual real-life emotions—as it happens in reflex-driven animal vocalizations (Seyfarth and Cheney, 2017). Their acoustic form is shaped by the physiological impact of emotion on the vocal organs plus Pavlovian-style priming.

Semiosis originates in an ongoing interaction between signalers and receivers within the reference-framework of the *same* environment—forging communication rules through the dialectics of ritualization and devaluation (Wiley, 1983). Ritualized signals establish conventions via encoding/decoding interaction between the acquainted individuals. Once established, convention becomes "devalued"—abused by "bluffing calls" of the unacquainted signalers trying to take advantage of the established reactions of the receivers. Increase of dishonest signaling causes the signaler to substitute the signal or modulate it along a single acoustic dimension until an "evolutionary stable strategy" is formed, marking a stationary equilibrium within the population—which ultimately fixes the convention (Maynard-Smith, 1976). Here, "signaling efficacy" obtains its formative power: as natural selection optimizes a signal to support the signaler's visual display, successful decoding starts relying on whatever the receiver finds most comfortable to detect, discriminate, and remember (Guilford and Dawkins, 1991). Together, strategic design and efficacy determine the ultimate structure of a signal.

The road from animal call to musical phrase goes through the ritualization of innate physiological and behavioral cues that animals use to exchange information (Maynard-Smith and Harper, 2003)[20]. Ritualized signals differ from cues by being more conspicuous, redundant, stereotypical, and containing alerting components (p. 72). Nevertheless, they remain "concrete" (bound to a single context) like cues (Fitch, 2010, 184) and unlike "transposable" music. For ritualized signal to evolve into musical phrase, its meaningful features must be abstracted to become non-signal-specific and form an AE of TO—a conventional dimension of gradient change along some axis.

The end result of such abstraction is the multifactorial nature of music communication (**Figure 1**): each emotional/motivational state is represented not by a *dedicated signal* but by the configuration of *numerous AEs* (Juslin, 2005). Conventional musical notation is poorly suited for incremental representation of AEs other than rough indications for melody/harmony, rhythm/meter, and form. Waveforms display rhythm and dynamics in finer detail, but miss other AEs. Spectrograms decently represent melody, rhythm, articulation, register, harmonicity, and dynamics, but miss harmony, tempo, meter, and texture. This necessitates the use of a special notation—such as *prosogram*, developed by Mertens (2004) for analyzing speech. Although applicable to monophonic vocal music in visualizing pitch, rhythm, articulation, dynamics, harmonicity, and register, prosogram ignores harmony, tempo, meter, texture, and form. To overcome these limitations, I propose a similar approach to music—"*musogram*[21]." Its advantages over conventional notation in capturing 11 AEs are demonstrated in the simplest case of classical music (**Figure 1**). It introduces the conventions, necessary to read the upcoming figures.

Multifactorial visualization reveals the expressive contribution of all AEs. Each AE features structural patterns representing specific emotional states across cultures, genres, and styles—at least for basic emotions (**Table 3**)[22]. Configuration of such patterns distinguishes one emotional expression from another. If multiple expressions share the same pattern of AE (e.g., legato characterizes both sadness and tenderness), the combination of a few aspects (e.g., "articulation + meter") differentiates them.

Multifactorial particulate semiosis shapes musical signs—each AE features SFTO, which enables "natural selection" for the most effectively communicated expressions. AC can be multifactorial but lacks particulate semiosis. Verbal semiosis is particulate but mostly unifactorial: phonetic organization is its primary source[23].

Basic emotions can be recognized across musical cultures (Mohn et al., 2010) and can be acoustically described (Eerola and Vuoskoski, 2013). Therefore, at least some of their musical markers share biological roots with mammalian ACs (Zimmermann et al., 2013). The birth of SFTO is trackable by comparing the multi-cultural markers of typical musical expressions of basic emotions to equivalent AC expressions and by inferring their differences and commonalities (**Table 4**). Common traits indicate music's inheritance from ACs, whereas contrasting traits—innovations brought about by cultural evolution.

Music and ACs have in common only regularity/irregularity and articulation. They both find a perfect match between human music and AC (5 out of 5 emotional states). The next closest match (4 out of 5) is "harmonicity." That is why these two aspects of TO (articulation and harmonicity) must be the most ancient, possibly retained from the pre-human times. In contrary, "register" shows a nearly perfect

[20]Maynard-Smith and Harper give an example of such ritualized physiological cues as thermoregulation that causes animals to raise their feathers/hair to reduce body temperature, heightened in social interaction—which makes an animal appear larger and promotes dishonest signaling of increased body size in instances of confrontation (p. 68). Other physiological cues are respiration, urination/defecation, pupil dilation, and yawning (p. 69). The ritualized behavioral cues include "intention to move" which signals the beginning of a significant action (a bird taking a few false starts before flying), "protective movement," and "displacement behavior" (p. 70).

[21]For thorough explanation of the visual representation of the multifactorial organization of music, a way of its quantification, and its difference from the prosogram approach by Mertens, see Appendix 1 "A New Method of Modal Multifactorial Analysis of Tonal Organization in Music" in **Supplementary Material**.

[22]Musicological literature identifies many more structural patterns of different AEs than the patterns listed in **Table 3**—and their semantic references include many more affective states than merely five basic emotions. Much of this information is dispersed in the treatises on music theory, some of which are cited in the beginning of this paper. There are very few books that list such structural patterns in a manner of the 18th century treatises of "musical lexicon" (Cooke, 1959; Mattheson and Harriss, 1981; Bartel, 1997; McCreless, 2002; Vashkevich, 2006). However, only isolated patches of such literature have attracted attention of psychoacousticians and received experimental trial (Kaminska and Woolf, 2000). For this reason, the metareviews on research in "musical emotions" tend to focus exclusively on 5 basic emotions.

[23]Although tempo, rhythm, prosodic contours, and registers contribute meaningful motivational and attitudinal information to verbal communication, by no means can they be regarded as its primary semiotic aspects. Without knowing the lexic meaning of words of a particulate language, inferred from phonetic structures of auditioned speech, no adequate understanding of that speech is possible. This is in polar opposition to musical semiosis, where tempo, rhythm, melodic contour, and register directly convey the most important information, whereas keeping the referential meaning optional.

mismatch, testifying that humans cardinally reorganized the use of registers in music. The rest of the AEs display mixed results. If to generalize by emotional states rather than by expressive aspects, then none of the emotions display a full match or a full mismatch. Evidently, coding of emotions in human music has developed its own proprietary acoustic attributes. This confirms that ACs are mostly conspecific. Heterospecific[24] generalities support only a rough distinction between "positive" versus "negative" emotions (Snowdon et al., 2015). Human communication inherits from ACs just 2 general semiotic oppositions: (1) positive/negative affectation and (2) low/high intensity of an affective state (Brudzynski, 2013). High-intensity "strong emotions" (Grewe et al., 2005) have evolved into chill-like experiences of music—in contradistinction to the "mundane" use of language (Silvia and Nusbaum, 2011). However, "strong emotions" *per se* could not support musical semiosis because the stimulus-response relationship between chill and music structure has not been experimentally reproducible—music chills seem to occur intermittently (Altenmüller et al., 2013a).

Both incremental and gradual changes in multiple AEs (**Table 1**) are peculiar to human music, whereas holistic tempo, dynamics, rhythm, and melodic contours are mutual for music and ACs. Musical meter, articulation, and harmony are also traceable to, respectively, ACs' regularity/irregularity, pausing/continuing, and periodicity/harshness.

However, the cross-examination of TO in expression of 5 basic emotions in music versus ACs reveals that many AE's patterns are unique to music (**Table 5**). Moreover, humans completely invert the acoustic characteristics of animal's affective states:

- Ascending/descending pitch (anger-tenderness),
- Fast/slow tempo (happiness-tenderness),
- Soft/loud dynamics (happiness-fear),
- High/low register (happiness/sadness-anger/fear),
- Harmonicity/inharmonicity (tenderness-anger).

This indicates massive remapping of the instinctive vocal encoding of affective states, achieved throughout the cultural evolution of Homo.

What could have caused such changes?

For many AEs, their cultural origin is obvious: metric pulses usually break into a default binary pulse (Potter et al., 2009), following the left/right paradigm instituted by bipedalism (London, 2004). Rubato patterns (ritenuto/accelerando) also relate to bipedal locomotion (Honing, 2003), so as tempo which is synchronizable to gait or heartbeat (Fraisse, 1982). Melodic intervals follow another locomotive paradigm of stepping/leaping (Nikolsky, 2015b)—each successive tone either "stands" (unison), "steps" (2nds and fast 3rds), or "leaps" (>3rd)—unlike harmonic intervals that are factored by consonance/dissonance relations (a much later historic semiotic development). Articulation grouping relies on yet another biological factor—the breathing cycle (Alekseyev, 1976, 130). Taking a breath terminates a phrase, imposing a "clausal structure" on the melody (Fenk-Oczlon and Fenk, 2009b). The "breath group" prototypes the "articulation group" via a "breathing pulsation" (Etzel et al., 2006). Noteworthy, breathing pulse takes over metric control in ametric forms of music-making (Wallin, 1983). Locomotive and respiratory AEs must have formed long before Homo.

The rhythmic aspect of music possibly emerged from the quantification of verbal rhyming, following the language development (Kharlap, 1972)[25]. Melodic contours also relate to verbal prosody. The timeline of language formation remains controversial: the "saltational" scenario regards language as a sudden mutation 50–100 kya, whereas the "gradual" scenario qualifies it as part of evolution throughout millions of years (Hillert, 2015). Paleoneurology points to the Middle Pleistocene as a birthtime of language (Quam et al., 2017). Since musical rhythm and melodic contours rely on fine vocal control, their addition to TO must have followed the accumulation of extensive lexic vocabulary within a phonological organization of language (Tallerman, 2013). This ties the emergence of multifactorial TO (which is hardly possible without engaging melodic contour and rhythm) to Homo sapiens and the Upper Paleolithic, as indicated by the proliferation of bone "flutes." During 1995–2009, over 120 bone pipes were recovered across Europe, dated 36–30 kya and concentrated up to 3 "flutes" per cave (Conard et al., 2009). Evidently, melodic music suddenly became popular in the Aurignacian.

Discreteness of pitch is evident in the construction of Paleolithic "flutes": holes are drilled in particular spots in order to generate sound of a particular pitch, and there is evidence of common patterns in the intervallic distances between the placement of the holes, suggestive of the commonality of certain melodic intervals in Aurignacian music-making (Nikolsky, 2015b, Appendix II). Discreteness of pitch was very likely to have been accompanied with the discreteness of rhythm, since stressing a pitch as a rule relies on extending its time-value relative to other pitches. Pitch hierarchy is supported by rhythmic contrasts between shorter timing of modally insignificant pitch-classes as well as longer timing of modally important pitch-classes (Krumhansl, 1990).

However, Aurignacian music most certainly lacked SFTO—semiotization of rhythm and directionality requires an extensive period of exploration. This is obvious in the acquisition of musical skills throughout infancy: infants babble—engage in meaningless play with melodic contours—before learning to compose musically *expressive* vocalizations (Moog, 1976; Dowling, 1984; Swanwick et al., 1986; Holahan, 1987; Hargreaves,

[24]The opposition of conspecific and heterospecific distribution of acoustic features that characterize the vocal expression of a particular affective state in AC allows a researcher to identify those patterns of AEs that match cross-cultural features of corresponding affective states in "musical emotions" of human music. The patterns of expression that are present across multiple animal species are more likely to form the equivalents of "universal" traits of human "musical emotions" than those patterns that are found only within the very same animal species.

[25]However, the idea of rhyming seems to have a precursor in ACs. Thus, humpback whales match the constituent syllables in some of their songs (Payne, 2001). A similar organization was noticed in mockingbird songs (Thompson et al., 2000). Its underlying cause is perhaps simplification of memorizing a complex song. Yet another cause could be the employment of repetition of a particular syllable in a song for a certain number of times as a conspecific marker for certain bird species (Fitch, 2010, 183). Hearing such birdsongs might have prompted humans to invent rhyming.

**TABLE 3 |** The configuration of structural patterns for each AE, typically used to express five basic emotions.

| Acoustic domains | Aspects of expression in music | Range of the expression of an aspect | Happiness' acoustic markers | Sadness' acoustic markers | Anger's (aggression) acoustic markers | Fear's (anxiety) acoustic markers | Tenderness' (love) acoustic markers |
|---|---|---|---|---|---|---|---|
| 1. Frequency | 1. Melodic pitch (consecutive "linear") | High/low relation of tones | Prevalence of ascending contour within a diverse set of contours, wide ambitus, leaps, sharp zigzags, and sharpened intonation | Prevalence of descending and wave-like contours, narrow ambitus, mainly stepwise motion, flat and sliding down intonation | Prevalence of ascending contour, with little diversity of other contours, frequent leaps with sharpened melodic contours and tendency to short motifs | Prevalence of ascending contour, of little variation, wide ambitus, many interruptions, frequent leaps, including extreme, use of angular and wave-like intonations | Fairly narrow ambitus, prevalent steps with occasional leaps, rising intonation, wave-like shapes |
| | 2. Harmonic pitch (concurrent "vertical") | Concordant/ discordant combination of tones | Major, diatonic, prevalence of medium size perfect intervals of fourth and fifth | Minor, chromatic, dissonance, prevalence of small intervals | Minor, chromatic, strong dissonance (up to atonal), large intervals, esp. major seventh, augmented fourth | Minor, strong dissonance, diverse intervals | Major, diatonic, general consonance |
| | 3. Form (thematicism) | Sameness/diversity | Relative simplicity | High complexity | High complexity | Relative complexity | Quite low complexity |
| 2. Time | 4. Tempo | Fast/slow metric pulse | Mostly fast, with very restricted use of rubato | Slow, with strong rubato and prevalence of ritenuto | Fast, with minimal rubato and general tendency to use accelerando | Fast, with strong rubato and many abrupt changes | Slow and moderate, strong rubato but no abrupt changes |
| | 5. Rhythm | Short/long relation of tones | Sharp contrasts of tones, yet smooth succession of groups of tones | Smoothened contrasts of tones, yet firm patterns, many long tones | Very sharp contrasts of tones, complex patterns with sudden changes, many short tones | Abrupt changes of tones and rhythmic groups, with overall diversity in rhythm, many short tones | Smoothened contrasts of tones, yet with rhythmic diversity, many long tones |
| | 6. Meter | Short/long periodicity of stressed beats | Strong regularity with minimal deviations | Tendency to irregularity | Tendency to syncopation and irregularity | Pronounced irregularity and variability | Strong regularity with moderate variability |
| | 7. Articulation | Styles of attaching/detaching of successive tones | Prevalence of staccato, with overall great diversity of styles | Prevalence of legato, with little diversity of styles, many pauses | Prevalence of staccato, with moderate diversity of styles, occasional legato | Prevalence of staccato (stressed), with great diversity of styles, many pauses | Prevalence of legato, with little diversity of styles, many pauses |
| 3. Amplitude | 8. Dynamics | Loud/soft relation of tones | Prevalence of loud and medium loud, with limited crescendo and diminuendo | Prevalence of soft and medium soft, with medium crescendo and diminuendo | Prevalence of very loud, with very little dynamic change, accents tend to fall on unstable tones | Prevalence of soft and medium soft, yet with diverse dynamic changes, mostly abrupt | Prevalence of medium soft, few dynamic changes, accents tend to fall on stable tones |
| 4. Timbre | 9. Register | Relation of homogenous groups of tones in their tonal quality | Prevalence of bright register, raised singing formant, brightness | Prevalence of bright register, low singing formant, dullness | Bright register with little changes, raised singing formant, harshness | Prevalence of bright register with abrupt registral changes, general mellowness | Prevalence of dark register, lowered singing formant, general mellowness |
| | 10. Harmonicity, attack and vibrato | Periodic/non-periodic spectral content | Harmonic richness, fast attack, medium vibrato with mid-fast rate | Harmonic scarcity, slow attack, small vibrato range with slow rate | Harmonic richness, much spectral noise, fast attack and decay, large vibrato range, mid-fast rate | Contrasts of harmonic richness and scarcity; gentle attacks; small irregular vibrato with fast rate | Harmonically moderate, slow attacks, small vibrato range with mid-fast rate |

*This table is compiled based on a number of meta-reviews of experimental research on emotional responses to listening to music (Gabrielsson and Lindström, 2001; Gabrielsson and Juslin, 2003; Juslin and Laukka, 2003; Juslin, 2005). The data is categorized according to the musicological nomenclature: all acoustic attributes are broken into 10 AEs across 4 acoustic domains. The aspect of texture is missing, because it was not controlled for in the experimental studies of the acoustic structural patterns that characterize "musical emotions". The aspect of harmonicity constitutes an organic part of the aspect of instrumentation, listed in the beginning of this paper. This potentially confusing mismatch occurs as a result of the discrepancy in musicological and psychoacoustic scholarships: as a rule, musicians are ignorant of harmonicity, while psychoacousticians are ignorant of instrumentation. Harmonicity can be defined as the extent to which the spectrum of a complex tone is made of its component frequencies that are integer multiples of its fundamental frequency (FF). This is usually measured as the ratio of harmonics to noise. Slow attack and great vibrato generally tend to reduce harmonicity in a monophonic tone.*

**TABLE 4 |** Acoustic attributes of typical animal vocalizations used by different species to display their affective state, grouped according to AEs of human music.

| Acoustic domain | AEs of calls | Range of an AE | Pleasure features | Dissatisfaction features | Aggression features | Alarm/disturbance features | Appeasing features |
|---|---|---|---|---|---|---|---|
| 1. Frequency | 1. Pitch (melodic contour) | High/low relation of tones in a call | Great variability of frequency modulations | Descending contour accompanies the increase in emotional intensity | Prevalence of descending contour, wide ambitus; yet use of rising pitch, few modulations; leaps accompany the rise in emotional intensity | Prevalence of ascending contour, wave-like and flat melodic shapes, great modulations; leaps accompany the rise in emotional intensity | Prevalence of ascending contour, narrow ambitus, great variability of modulations |
| | 2. Harmony | n/a | n/a | n/a | n/a | n/a | n/a |
| | 3. Form (complexity) | Compact/expanded call structure | n/a | Long and short calls, longer as intensity increases | Long calls, but generally shorter than alarm calls | Longer calls | Longest calls |
| 2. Time | 4. Tempo | Fast/slow pace of calling | Slow | The slowest, but speeding up as emotional intensity increases | Fast, speeding up more as emotional intensity increases | Fast, speeding up more as emotional intensity increases | Fast |
| | 5. Rhythm | Short/long duration of tones in a call | Generally short tones | Generally long tones, often repeated | Prevalence of long tones, longer as emotional intensity increases | Prevalence of short tones, longer as emotional intensity increases | Short tones, very repetitive |
| | 6. Meter (regularity) | Periodicity of calls | Regularity | Irregularity | Irregularity | Irregularity | Regularity |
| | 7. Articulation | Attaching/detaching of tones in a call | n/a | Distinct pausing | Abrupt ending of a call, staccato, shortened pauses | Smoother ending of a call, staccato, shortened pauses | Smoother ending of a call, legato, distinct pausing |
| 3. Amplitude | 8. Dynamics | Loud/soft relation of tones in a call | Generally soft | Medium soft, the increase in amplitude accompanying the increase in emotional intensity | Very loud, with abrupt onset and offset, the increase in amplitude accompanying the increase in emotional intensity | Medium loud, without abrupt onset and offset, greater increase in amplitude as emotional intensity increases | Soft, rhythmic sequences of amplitude modulation |
| 4. Timbre | 9. Register | Bright/dark relation of homogenous groups of tones in a call | Dark register (e.g., purrs and growls) | Dark register (e.g., meows), no change as emotional intensity increases | Very dark register, but increase in intensity engages brighter register (more so than in alarm calls), wide range | Mostly dark register, sometimes very dark, but increase in intensity engages brighter register | More of dark register (in some species bright), no change as emotional intensity increases |
| | 10. Harmonicity | Periodic/non-periodic spectral content of tones in a call | n/a | Tendency to use pure tones (whistle-like or flute-like) | Harsh, non-harmonious, noisy, broadband (growl, bark), harsher as emotional intensity increases | Mostly pure tone (whistle-like), can be noisy and broadband; both, harsher and/or richer in harmonics as intensity increases | Harmonic richness (grunt), mostly noisy, but much less harmonious as intensity increases (coo) |

*The data for this table is compiled from numerous meta-reviews (Morton, 1977; Peters, 1984; August and Anderson, 1987; Snowdon, 2003; Briefer, 2012; Altenmüller et al., 2013; Zimmermann et al., 2013; Snowdon et al., 2015). According to the classification scheme of Brudzynski (2013), human and animal affective states are equated in the following ways: human "happiness" is equated to animal "pleasure" (satisfaction), human "sadness"—to animal "dissatisfaction" (social isolation from a bonded party), human "anger"—to animal "aggression" (agonistic behavior, conflict with display of threat or combat), human "fear"—to animal "alarm/disturbance" (anxiety at the presence of threat or intimidation by a novel environment), human "tenderness/love"—to animal "appeasing" (affiliation—physical contact without agonistic behavior, e.g., grooming, and play). Those acoustic features that agree between human and animal expressions of the same affective state are marked blue, whereas the disagreeing features—red. Features that are not covered in research literature are marked "n/a." The aspect of "harmony" is clearly not applicable to animal vocalization. The aspect of "form" bears only distant relation to "musical form": AC's compactness loosely corresponds to simplicity of structure, whereas lengthiness—to complexity. Aspect of "meter" also finds only partial correspondence in regularity or irregularity of call units in the AC bouts. The timbral coloration is reflected by the aspect of "harmonicity" rather than "instrumentation" that manages timbre in human music.*

**TABLE 5 |** The acoustic attributes of typical expression of 5 basic emotions in human music that find no correspondences in animal communication (based on **Tables 3**, **4**).

| Acoustic domains | AEs of music | Range of an aspect | Happiness features | Sadness features | Anger (aggression) features | Fear (anxiety) features | Tenderness (love) features |
|---|---|---|---|---|---|---|---|
| 1. Frequency | 1. Melodic pitch (consecutive "linear") | high/low relation of tones | prevalence of ascension in the overall diverse contours, leaps, zigzags, sharpened intonation, wide ambitus | smooth contours, mainly steps, wave-like shapes, flat and falling intonation, narrow ambitus | prevalence of ascending contour, with sharpened shape of melodic contours | little variation in contours, interruptions, use of angular shapes, wide ambitus | prevalence of descending contour, mostly steps with occasional leaps, rising intonation, wave-like shapes |
| | 2. Harmonic pitch (concurrent "vertical") | concordant/discordant combination of tones | major, diatonic, prevalence of perfect 4th and 5th | minor, chromatic, dissonance, mostly small intervals | minor, chromatic, strong dissonance, wider intervals (major 7th, aug. 4th) | minor, strong dissonance | major, diatonic, general consonance |
| | 3. Form (complexity) | sameness (simplicity)/diversity (complexity) | relative simplicity | high complexity | n/a | n/a | low complexity |
| 2. Time | 4. Tempo | fast/slow metric pulse | fast, with very restricted rubato | strong rubato and prevalence of ritenuto | minimal use of rubato | strong rubato and many tempo changes | slow and moderate, with moderate rubato |
| | 5. Rhythm | short/long relation of tones | sharp contrasts of tones, yet smooth succession of groups | smoothened contrasts of tones due to frequent use of rubato | very sharp contrasts of tones, complex patterns with sudden changes | abrupt changes of tones and groups, general prevalence of diversity | smoothened contrasts of tones due to rubato use |
| | 6. Meter | periodicity of beat grouping | minimal deviations | n/a | tendency to syncopation | pronounced variability | moderate variability |
| | 7. Articulation | styles of attaching/detaching of successive tones | prevalence of staccato, great diversity of styles | prevalence of legato, with little diversity of styles | moderate diversity of articulation, occasional legato | stressed staccato (marcato), great stylistic diversity, many pauses | little diversity of styles |
| 3. Amplitude | 8. Dynamics | loud/soft relation of tones | prevalence of loud and medium loud, with limited crescendo and diminuendo | prevalence of soft, with medium crescendo and diminuendo | mostly loud, with very little dynamic changes, accents tend to fall on unstable tones | prevalence of soft and medium soft, yet with diverse dynamic changes, mostly abrupt | few dynamic changes, accents tend to fall on stable tones |
| 4. Timbre | 9. Register | bright/dark relation of homogenous groups of tones | prevalence of bright register, raised singing formant, brightness | prevalence of bright register, lowered singing formant, dullness | prevalence of bright register, with little change, raised singing formant, harshness | prevalence of bright register with abrupt registral changes, mellowness | prevalence of dark register, lowered singing formant, mellowness |
| | 10. Harmonicity, attack and vibrato | periodic/non-periodic spectral content of tones | harmonic richness, fast attack, medium vibrato with mid-fast rate | slow attack, little vibrato with a slow rate | harmonic richness, fast attack and decay, large vibrato with a mid-fast rate | gentle attacks; little and irregular vibrato with a fast rate | slow attacks, little vibrato with a mid-fast rate |

*These attributes constitute a stock of TO features developed in the process of evolution of human music from hominin musilanguage. This includes changes in vertical harmony, in metric pulse, and in complexity of musical form; contrasts in melodic contour, in directionality of melodic intervals (sharpening for ascending, flattening for descending dyads), and in thematic material; diversity of rhythm, articulation and tempo; and ambitus size. Animal vocalizations do not seem to engage these categories in meaningful differentiation of calls.*

1996). Most children pass through a music-babbling stage when 12–18 months old (Gembris, 2006). Universality of babbling suggests the universality of prolonged sensorimotor trials in music-making before semiotic rules are formed. Babbling abstracts melodic directions and intervals, allowing an infant to master particulate semiosis. Similarly, early humans had to long experiment with meaningless melodic play for the SFTO conventions to emerge.

## CROSS-CULTURAL "SCRIPTS" IN THE FORMATION OF SEMIOTICALLY FUNCTIONAL TONAL ORGANIZATION

Tool-making technologies (Ambrose, 2001) and "social scripts"—i.e., fixed generalized patterns of social behavior (Aiello, 1998)—most likely served as syntax precursors by providing explicit models for combining numerous elements into a structured sequence (Wildgen, 2004). Paleolithic proxies for syntactical language include composite tools (Ambrose, 2010), fire (Brown et al., 2009), knot-making (Camps and Uriagereka, 2006), cooperative hunting (Chase, 2006, 52), symbolic behaviors (Mcbrearty and Brooks, 2000), and burials (Mellars, 2004). The same proxies apply to syntax-related features of musical TO. All the AEs of music listed above (perhaps, except harmonicity) are engaged in the syntactic organization of music. Phrasal ends are usually marked by descending pitch, lower register, more concordant harmony, slowing of tempo, longer rhythmic value(s) placed on metrically strong time, reduction in loudness, and clear caesuras in articulation which separate the end of one formal unit (phrase, sentence) from the beginning of the following unit. In addition, there is evidence of a link between structures of tonal and social organization in indigenous societies (Blacking, 1967; Davidson, 1970; Lomax, 1977; Berliner, 1993; Arom and Voisin, 1997; Kubik, 1999)—which indicates that social structures might have also served as proxies for music syntax.

Making bone "flutes" was extremely tedious, demanding skills and expertise (Münzel and Conard, 2009). Why to invest into a "pitch toy" rather than to merely vocalize?

Cave-inhabitants must have supported flute-makers in the same way as they supported cave-artists—their exquisite labor required narrow specialization, precluding participation in hunting/gathering. In animistic ideology, depictions linked hunters to prey, providing means to benefit the outcome of hunting (Hauser, 1999, 1–4). Magic—not aesthetics—governed rock art, turning depiction into a shamanic occupation[26]. Shamanic music resembles shamanic depiction by cross-linking the signified to the signifier (Hubbard, 2003). In northern shamanic traditions, both melodic and pictorial contours are believed to affect the corresponding real objects (Novik, 2004, 67–85). Archeological evidence also links most resonant locations in caves with rock art in Paleolithic sites, suggesting the combined ritualistic use of images and music (Reznikoff, 2008; Morley, 2013; Mills, 2016). Hence, a Paleolithic "flute" was most likely a talisman used in rituals (Marshack, 1990). Its manufacturing from the bone of a particular animal (Wyatt, 2016) must have carried more significance for Aurignacians than the pitches it produced.

For melodic semiosis to occur, rhythm and directionality must first be abstracted into AEs. Abstraction of directionality probably followed rhythm: *salience of the melodic direction depends on rhythmic values, but not vice versa*. Tracking the melodic contour within the tonal "grid" constitutes the backbone of melodic organization (Deutsch, 2013), just like tracking the rhythmic grouping within the metric grid supports the temporal organization (Large, 2008). Reference to tonal hierarchy interferes with rhythmo-metric perception by biasing the attention toward pitch (Prince et al., 2009). Their conflict indicates that users of non-Western music discriminate rhythmo-meter better than users of Western tonality (which agrees with the observations of ethnomusicologists). This suggests that *frequency reference-frame* emerged later than *rhythmo-metric*.

Developmentally, acquisition of rhythmic hearing usually precedes melodic hearing (Shatkovsky, 1986). Infants seem to acquire rhythm-discrimination skills earlier than pitch-discrimination (Trehub and Hannon, 2006)[27]. The perceptual foundations of rhythm/meter are manifested just a few days after birth, as a part of developmentally crucial rhythmic interaction between infants and caregivers, occurring spontaneously and requiring little experience—reflecting its evolutionary importance for bonding (Trainor and Hannon, 2013). In verbal acquisition, rhythm too obtains semantic functionality earlier than prosodic contour (Shvachkin, 1948). According to the vast data collected through administration of early musical education in USSR, rhythmic hearing lays the foundation for vocal musical skills—followed by learning to reproduce melodic contours (Kirnarskaya et al., 2003, 168–170). Impressions that not only rhythm can influence melodic perception by directing the attention to longer tones, but that melodic features carry the reverse influence onto rhythm, are based on the misnomer between rhythm and meter (McAuley, 2010). Melodic intervals, contours, and "tonal accents" help to infer meter, but play no major role in identification of rhythmic values. On the contrary, judgments of melodic similarities are significantly affected by rhythm, especially in folk music (Eerola et al., 2001)[28]. Even for experienced Western musicians the distinction between

[26]Thus, newer paintings often covered the older ones: hiding the underlying image did not matter—once painted, an image was "brought to life," and stayed "alive," even if masked—just as a person who disappears from our sight does not die (Uspensky, 1995, 173–181).

[27]The earliest age when infants show the ability to recognize changes in pitch contour is 5 months (Chang and Trehub, 1977). Majority of studies demonstrate such capacities in older children, 6 months and up (Trainor and Hannon, 2013). The ability to recognize changes in rhythmic values of a familiar music seems to emerge quite earlier—at 2 months of age (Demany et al., 1977).

[28]Metricality, along with tonality, influence primarily the Western musicians: non-musicians process melodic contours mostly according to the distribution of longer rhythmic values (Monahan et al., 1987). Non-trained listeners simply cannot ignore rhythm—as it governs their melodic recognition (Jones and Ralston, 1991). Majority of young and inexperienced listeners at first parse melody by rhythm and only then by pitch contour and mode (Halpern et al., 1998). Tempo/rhythm descriptors are much more prevalent in listeners' judgments of thematic similarity than of pitch contour (Addessi and Caterina, 2000; McAdams, 2004).

rhythms is more salient than the distinction between pitches (Monahan and Carterette, 1985)[29].

Important Upper Paleolithic cultural proxies promote the abstraction of rhythm—not of melodic contour. Metric pulse is transposable from bipedal gait into such a common Paleolithic activity as stone-knapping. Each knapper prefers his own tempo and rhythm (Whittaker, 1994, 81)—quite similar to individual gait preferences (Whittle, 2007). Knappers' heartbeat provides a metric reference (Zubrow and Blake, 2006). Two knappers might have accidentally discovered the expressive capacity of rhythm through their entrainment, thereby forming the world's first musical instrument (Montagu, 2004). Group "musical" knapping was observed amongst Aboriginal women in Queensland (Duncan-Kemp, 1952, 27). Rock slides and gongs are drummed across the globe in rituals related to fertility cults (Fagg, 1997, 38). The ritualistic context provides feeling of contentment or awe, abstractable into a semantic value for the knapping/grinding sound, turning its rhythm into a sign—and the archeological evidence for collective stone-knapping is present in Neolithic sites at Sanganakallu-Kupgal, India (Boivin et al., 2007). Even earlier, stationary lithophones were drummed in Solutrean-Magdalenian caves (pecked rock surfaces were found in Africa)—suggestive of the existence of portable lithophones (Blake, 2011). The weird-sounding cave echo might have prompted specific affective connotations (Cross and Watson, 2006).

Unlike rhythm, *pitch directionality* finds no proxies in the Paleolithic[30]. A set of meaningful pitch contours could have originated in verbal prosody, but paleolinguists connect the development of the fully phonemicized semantic languages to population growth *after* the Last Glacial Maximum (Robb, 1993). Deeply social, language is imperative for accumulation of knowledge, which depends on population density to avoid "bottlenecks" due to climate changes and extinctions. Cultural evolution stabilized only after 50 kya—most certainly, because of the advancement of language (Klein, 2009). In all the prehistory, the transition to Holocene stands out as the grand leap in innovation, called to subsist an ever-growing population (Richerson et al., 2009). Powell et al. (2009) developed a demic model to estimate the critical population density capable of sustaining the innovation growth to offset the innovation loss: for Europe it was 45 kya. Prior to 20 kya, prehistory consisted of a chain of major discontinuities in cultural transmission (d'Errico and Stringer, 2011). Technically, the archeological concept of "culture" applies only starting from the Neolithic (Probst, 1991, 227).

The first archeological symbolic "culture" of pan-European scale is the Gravettian, whose common trans-European traits are both socio-economic and spiritual, with regional differences confined to the material techno-complex (Kozłowski, 2015). The continent-wide cultural unity is evident in the omnipresence of "Gravettian Venuses" over most of Europe (Soffer et al., 2000)[31]. Denser population turns language from means of inter-group cooperation that compensates for local ecological deficits into a life-long ethnic marker, akin to the cranial configuration (Robb, 1993). Personal ornaments in Gravettian burials manifest similar function of the "ethnic badge," differentiating age classes across the puberty threshold (Zilhão, 2014).

Social restructuring by ethnos and age hardly occurred without the involvement of music, closely affiliated with funeral and puberty rites. The Gravettian funerary practice strongly suggests the existence of burial rituals regulating the emotive interaction between the group's members, the dead, and the landscape as part of a greater ritual system, underpinned by cosmological beliefs (Pettitt, 2010). The remnant of such socio-eco-cosmological interconnection with TO, providing its semantic foundation, is the ancient doctrine of **ethos**[32]—renowned in Hellenic civilization (Mathiesen, 1984), but certainly much older (Farmer, 1965) and geographically wider (Manuel and Blum, 2011). The roots of ethos must lie in the Gravettian trans-European spiritual unity.

## CONTRIBUTION OF MULTI-DIMENSIONAL AND MULTI-EMOTIVE SEMIOSIS TO THE EVOLUTION OF MUSIC

Human melodic universals remap animals' universals. Animal anger is characterized by descending contour, whereas animal appeasing—by ascending contour. Music reverts the registers for happiness, sadness, fear, and anger from low to high. Why?

*Music contributes to the conservation of knowledge* by bonding social groups and incentivizing linguistic communication. This capacity came in play after the Younger Dryas (11 kya), when global warming enabled colonization of Eurasia. Widely dispersed populations created a few flexibly bounded

[29]Of course, the influence of rhythmic features on the judgment of melodic similarity is far from being simple and direct. Other factors, such as tempo and harmonization, can affect the extent of autonomy of temporal and frequency-related aspects of music (Prince, 2014).

[30]There are accounts of "tone-painting" where the contour of the hills is represented through the melodic contour in songs of indigenous hunter/gatherers of Northern hemisphere (Krushanov, 1987, 234) whose life style is comparable to that of Aurignacians. However, the idea of such representation most probably was inspired by the need in mnemonic aid in long-distance navigation during migrations with reindeer herds, which doubtfully existed earlier than a few thousand years ago (see the last chapter). Such tradition had chances to survive the ongoing extinctions in harsh climate only as a part of a reliable subsistence strategy for a fairly large population.

[31]Broad-scale technological clustering originated in the earlier Aurignacian tradition—attributed to the long-term influence of the ethnolinguistic variation. Forming of a continental culture during the Gravettian indicates the increased language contacts between different "clusters," establishing pan-European networks of informational exchange (Zilhão, 2014).

[32]The term "ethos" was coined in Archaic Greece, where it originally meant "custom," but by Classic times it obtained the meaning of a certain affective "character," associated with a particular musical melodic mode. "Ethos" embodied the consensus within a community as to which affective states would be generally "good" or "bad" for that community. The doctrine of "ethos" is closely related to the concept of "harmony of spheres," attributed by Hellenic sources to Pythagoras, who presumably learned it from Babylonians. The discussion of ethical value of this or that musical emotion and its suitability for astrological dispositions constituted an important part of public discourse in Ancient civilizations of Near and Far East, as well as Central Asia.

"social territories[33]," developing the dialect continuums by linkages among groups due to intermarriages during population shortfalls (Robb, 1993). Population growth and sedentism accompanied rapid neolithization, promoting ethnogenesis and thereafter fissioning language into language families as regional cultural differences cumulated (Robb, 1991). Such line of development benefited from the social bonds established by music.

The absence of music-like particulate emotional communication must be one of the reasons why chimpanzees do not accumulate cultural traditions. Some chimpanzees acquire a culture of tools but due to the lack of transposability and abstraction cannot transmit it (Whiten, 2011). However, it is music, not language, that engages reproduction, transposability, and abstraction of idiomatic patterns of each of its AEs.

Human remapping of pitch encoding most probably originates from the continuous practice of:

- **Frequent rotation between aesthetic emotions:** ACs prioritize negative emotions due to greater urgency of their triggers (August and Anderson, 1987). Human music is balanced between negative and positive expressions because of the mentalization of aesthetic emotions (Juslin, 2013). Expression of negative emotions can be pleasurable whenever it occurs in a non-threatening situation, is aesthetically appealing, and seems somehow useful or appropriate (Sachs et al., 2015). Thus, abstraction of emotions enables older children to learn to appreciate sad music (Schubert and McPherson, 2015), whereas at 5–7 months, infants overwhelmingly prefer happy to sad music (Nawrot, 2003). By 4 years, children start intentionally expressing positive and negative emotions in singing (Welch, 2006), distinguishing happy/sad and angry/fearful musics (Eerola and Vuoskoski, 2013). This line of development is also applicable to cultural evolution. In both cases, changes of musical emotions sharpen *contrasts in patterns of their musical expression*—resembling *phonemic oppositions in phonology*.
- **Multifactorial musical semiosis:** Zero- and first-order intentionality separates animal signals from second-order intentionality of humans (Seyfarth and Cheney, 2017). Although non-human primates can coordinate the produced signal with the listeners' response, modulating the acoustic features of their calls accordingly, modulation usually engages a single parameter—falling short of the complex multidimensional nature of emotional communication in verbal prosody and music (Filippi, 2016). Simultaneous interactive control over *multiple AEs* is peculiar to music alone. Thus, in expression of anger, prevalence of ascending contour and high register conveys physical strain, while the side-effect of their monotony is compensated by a diverse contrasting rhythm and spectral content, projecting agitation (**Table 2**). AC's anger does not engage such interaction. It conserves a *unifactorial* timbral quality[34] (**Table 3**).

All AEs differ in musical expression of love (**Figure 2**) and anger (**Figure 3**), as evident in musograms[35] of indigenous Siberian songs that Russian theorists believe to represent the earliest forms of TO (Alekseyev, 1976, 1986; Brodsky, 1976; Zemtsovsky, 1983; Mazepus, 1993; Mazepus and Galitskaya, 1997; Novik, 1999; Zabolotskaya, 2009; Dobzhanskaya, 2011, 2016; Nikolsky, 2015b; Sheikin, 2017, 2002).

Unlike the expression of love, anger engages a wider ambitus, greater leaps, contrasting registers, harsh timbres, loudness, shorter and richer rhythms, reduced regularity and tonal stability, increased tempo fluctuations, staccato articulation, and thematic complexity (**Figures 2**, **3**). However, gorillas express anger differently: "call-motifs" remain always isolated and slow-paced, featuring neither a clear melodic contour (due to its enormous bandwidth) nor rhythm (**Figure 4**).

If humans *consciously* manipulate numerous *learned* expressive parameters in music, animals *instinctively* "center" on a single biologically "*hard-wired*" parameter to reflect their emotional intensity. Human infants start their development at the same level where animal cubs start theirs, but quickly advance. Newborns employ just 2 vocalization types: negative and positive (Loewy, 1995). Cries of hunger, cold, distress come first as biological reflexes (Zeskind, 1985). However, the similarity of an infant's supralaryngeal vocal tract to that of the primate cub's does not stop the infants from trying to imitate his/her caretaker's vocalizations (Lieberman, 1985)[36]. Infant cries start varying in temporal and frequency characteristics as the infant ages (Papoušek and Papoušek, 1995). Loudness, timbre, register, attack speed, FM range, and harmonicity are progressively mastered as markers of different cry-types (Golub and Corwin, 1985). An infant builds a repertory of melodic contours assigned to specific situations and used as building blocks to inform the caretaker about his/her state and to receive a desired treatment (Wermke and Mende, 2009). Such ongoing two-ended communication lies at the heart of musicality (Trevarthen, 2019).

Call/cry-repertory building appears to be universal in human development (Wermke et al., 2007), very likely paralleling the phylogenetic evolution of music (Foster, 1994). Similarities between the structure and function of human and non-human vocalizations were discovered in crying, motherese, and babbling (Snowdon, 2003). Fluent switching from one cry-type to another, corroborated by the caretaker's response,

[33]Thus, Peter Bogucki counts as few as 14 Mesolithic "social territories"—i.e., regions differentiated by the material culture as manifested by archaeological evidence—spread out over the entirety of Western Europe during its transition from the Boreal to the Atlantic periods, c. 7500 kya (Bogucki, 1988, 41–46).

[34]It could be said that an animal "centers" (i.e., focuses) on a single aspect of vocal expression, conserving the extent of increase or decrease in intensity of the psycho-physiological state that is associated with that vocal expression. This is yet another parallel between AC and the vocalization of a sensorimotor human infant. This is in contrast to the ability of an adult human to simultaneously conserve multiple dimensions of changes in multiple AEs in music.

[35]For a comprehensive analysis of those musical examples that were selected for musograms in **Figures 2**, **3**, **4**, and **7**, see Appendix 2 "A Comparative Structural Analysis of Musograms" in **Supplementary Material**.

[36]Similarity of the anatomy of the supralaryngeal vocal tract of the human baby and the ancestors of Homo sapiens provides yet another justification for seeking the TO model of hypothetical Paleolithic music in the musical babbling of 1–2-year-old infants.

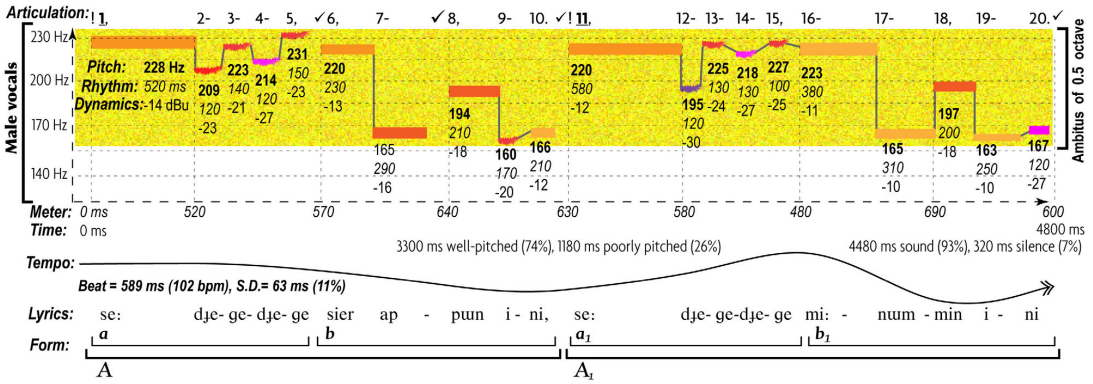



**FIGURE 2 |** Characteristic patterns of AEs in expression of love in a Yakut traditional lyrical song "Sae Dyige" (may be auditioned at http://chirb.it/sNegG1). By Juslin's (2005) classification this song fits the "love" music category—in agreement with its lyrics, describing how a woman is anticipating visits of her multiple lovers (Alekseyev and Nikolayeva, 1981, 86). The musogram follows the same conventions as **Figure 1**, with minor additions due to the less definite use of pitch in the purely vocal music. Tones of low spectral periodicity (noisy or spoken-like) are represented by fuzzy strips in contrast to high periodicity, represented by rectangular bars. The number under each pitch displays its frequency value in bold, its duration in italic, and its maximal amplitude (the highest value of any of its spectral constituents) in regular font. The lyrics are given in the phonetic transcription. There are two contrasting motifs: "a"—a sustained long anchor tone (tonic function), followed by rapid alternation of steps with rising intonation; and "b" —two descending intonations, the first of which leaps to the alternative anchor (dominant function to mark a cadence), while the second steps down and then gently rises. These two motifs make up a call-like phrase that is regularly repeated. Song is characterized by a narrow ambitus (half-octave), mid-low register, high harmonicity, low complexity, moderate tempo (102 bpm) with little rubato (11%), diverse rhythm (usage of four rhythmic values), regular meter, overwhelming legato (97%), and scarce dynamic changes. For more detailed discussion, see Appendix 2 "A Comparative Structural Analysis of Musograms."

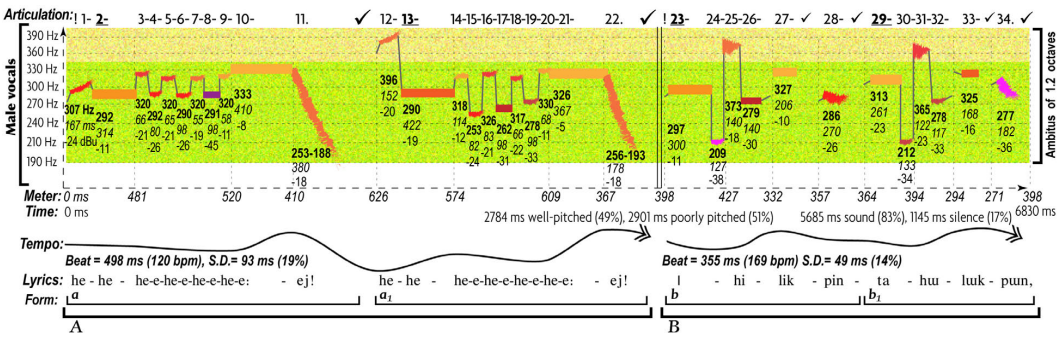



**FIGURE 3 |** Characteristic patterns of AEs in expression of anger in a song of the underworld virgin from the olonkho "Djiribina Djirilatta" (http://chirb.it/sCq02k). This excerpt from the traditional Yakut epic expresses anger of the evil sorcerer toward the heroine, challenging her to a fight (Alekseyev and Nikolayeva, 1981, 35). Structural descriptors of most aspects of this song fall in the category of "angry" music (Juslin, 2005). The acoustic markers of all AEs contrast those in **Figure 2**. The ambitus is over twice wider. There are two registers instead of one: low singing and high "shouting"), both are higher than **Figure 2**. The share of well-pitched sounds in the overall duration of music is reduced by 34%. The share of staccato articulation is increased (by 142% in the duration of silence and 40% in the number of pauses). Tones are overall shorter and 50% more diverse in time values, with contrasts between rhythmic groups. The tempo contains abrupt switches, the fastest of which is 66% faster and 73% more variable (rubato) than **Figure 2**. Intonations feature wide leaps, on average 70% wider than **Figure 2**. Thematically, the music is more diverse and complex, using two contrasting materials, "A" and "B" (**Figure 2** had only one). Timbre is harsh (a heightened larynx and intensified pressure).

prompts the cross-examination of the cries' acoustic parameters. The intensity of temporal expression usually matches pitch expression (frequent leaps require faster tempo to convey excitement and emergency—otherwise the caretaker is not "convinced" to respond urgently enough). Together, the projection of feedback and memorization/cross-relation of cry-types establish the acoustic oppositions between AEs of common musical emotions.

What diverts music from AC is the radical change in communication framework. Animals communicate "face-to-face" in situations that demand immediate action, which selects signals effective in expressing rapidly changing motivational states, with clear gradations in their intensity (Morton, 1977). Such signaling prioritizes ease of detection, speed of interpretation, signal's briefness, and a single salient gradient AE (Maynard-Smith, 1976). High redundancy and stereotypicity of selected signals often "fix" them (Simpson, 1997). This precludes combinability of AEs and calls, enabling "*dishonest*" calling.

Unlike animal calls, traditional indigenous music normally *never "lies"* (Nikolsky, 2016, Appendix III). A performer, as a rule, expresses emotions he actually feels—even when

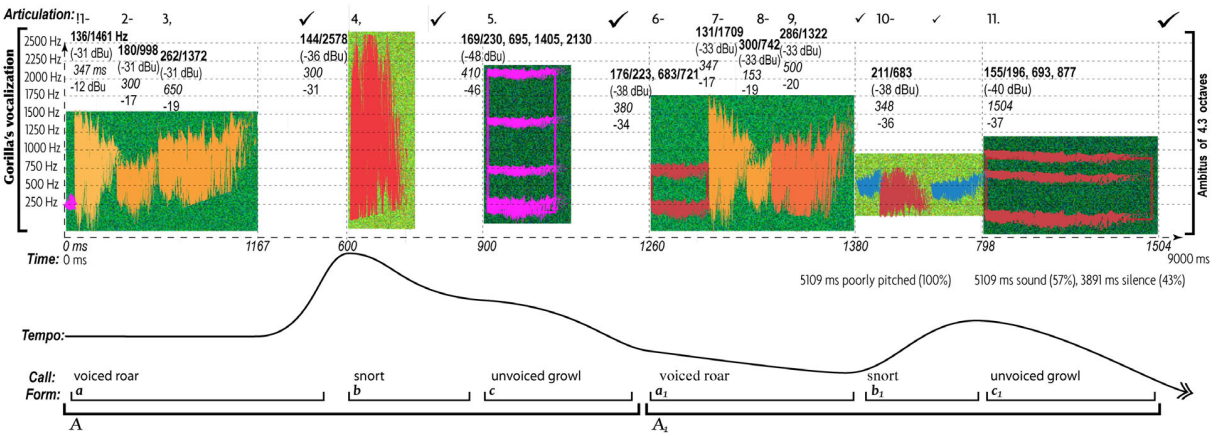



**FIGURE 4 |** Characteristic patterns of AEs in expression of anger in gorilla's calls (http://chirb.it/72g63y). Approaching primate's vocalizations with the same multifactorial analytical method as human music reveals important differences in TO. The most noticeable is complete absence of harmonious sounds with clear FF and legato articulation. The share of silence doubles: 43% (versus 17% of **Figure 3**). The form is simpler—no motifs conjoin into a phrase. Calls (voiced roar, non-voiced growl, and snort) remain detached except for a few instances of joining snort and growl together. The same disconnectedness characterizes all temporal AEs. The onset of each of the calls exposes a sort of an irregular pulse. However, the rate of this pulse is more than twice slower than the angry human music (**Figure 3**) and its deviation from a regular pulse is nearly twice greater—exceeding even the slow and flexible "loving music" (**Figure 2**). In essence, it would be accurate to characterize these vocalizations as rhythmically irregular, ametric, and undifferentiated in pitch. None of the calls generate a clear pitch contour due to their very broad band (up to 4.2 octaves). The calls' bandwidth was calculated by taking measurements of the frequency of that portion of the spectrum which stood out from the rest of the signal. Unlike music, gorilla's call-motifs do not break the ambitus into registers but timbrally recolor the entire ambitus for each of the calls, thereby increasing their separation.

impersonating an epic protagonist or a spirit, the singer becomes temporarily "possessed" by them (Novik, 2004, 272). "Putting on an act" is a prerogative of post-Renaissance Western classical performance tradition, and even there the performance canon demands "method-acting" to convince the audience in the realism of musical emotions (Nikolsky, 2015a)[37]. A non-western traditional song usually appears "westernized" to the indigenous audience when "acted out" formally (Zemtsovsky, 1983). Folk "cover-songs" necessarily engage the performer's "direct"—rather than "indirect" or "scripted" speech (Zemtsovsky, 1979)[38].

Insincerity and falsehood in musical expression did not present a critical issue prior to the 1760s (Charlton, 2009). They both attracted public discourse as a systemic aberration peculiar to a specific class of music (rather than a "defective" sample) only after the entertainment industry became institutionalized (Dahlhaus, 1989, 314). Rise of mass production made "emotional faking" a norm for commercial popular music—explicitly codified in Irving Berlin's composition standards (Suisman, 2009)[39]. So, music started as a decidedly "honest signal" (Levitin, 2009, 141–6) and only recently adopted "acting"—albeit, hardly enough to declare music fundamentally "dishonest[40]."

Jointly, multi-dimensionality of music and emotional contagion make lying difficult. Music always integrates listeners and performers, and this togetherness promotes sincerity. The particulate structure of musical semiosis effectively reveals dishonesty: at least some of AEs' insincere expressions are bound to contradict each other, prompting a resolving interpretation. But what in the cultural evolution could have spurred the inclination for aspect-matching?

[37]Demonstration of musical "method-acting" can be found in the video clip of Andrei Gavrilov performing Rachmaninov's Prelude in g minor, op.23, No.5 https://www.youtube.com/watch?v=T3AEfMMyH6A. Especially telling is the pianist's facial expressions, as he is getting up from the piano bench after completing his performance—he continues to remain in his "role."

[38]Some indigenous traditions have developed professional forms of musical art which require aesthetic evaluation (e.g., Tatar, Kazakh, Mongolian). However, they still fundamentally differ from Western classical music by not taking a musical work as a "script' created by the composer for the performer to adhere to (Zemtsovsky and Kunanbayeva, 2011). Only the Western musician is trained as part of his occupation to accurately "execute" the composer's script while being aware of the fictiveness of its emotional content. However, application of such treatment to a folk cover song is most likely to come across as fundamentally "inauthentic" and detrimental to the song (Moore, 2011).

[39]Berlin's rules Nos. 3, 6, and 9 call for the composer to please the consumer at the cost of insincerity: "the ideas and lyrics must suit *either a male or a female*, so both sexes want to *buy a song*," "music and lyrics *must* have to do with *things common to everyone*," and, most explicitly—"songwriter must look upon the song as a mere business, *not take music to heart*." Berlin's rules break away from the Western composer's "canon," established since the introduction of "musica reservata" in the 16th century (Meier and Dittmer, 1956). For this reason, Berlin's approach provoked criticism of the American popular music in toto, seen by connoisseurs of art music as a "sweet lie" sold (for profit) to the mass audience to replace music that is "truthful" yet unpleasant in revealing "social truth" (Adorno, 1942).

[40]As far as I know, Trehub (2008) remains the only scholar who believes that music, in general, operates by having the performer emotionally deceive the audience. Other scholars who point out that a professional performer can evoke emotions that he/she does not actually feel, realize that this discrepancy is possible only in music that segregates the listener, the performer, and the composer. *This solely happens in Western classical music*. And even within this tradition "deceiving" the audience is still regarded as a fault to be avoided. Noteworthy, Trehub did not respond to Juslin and Västfjäll's (2008) objection to her criticism.

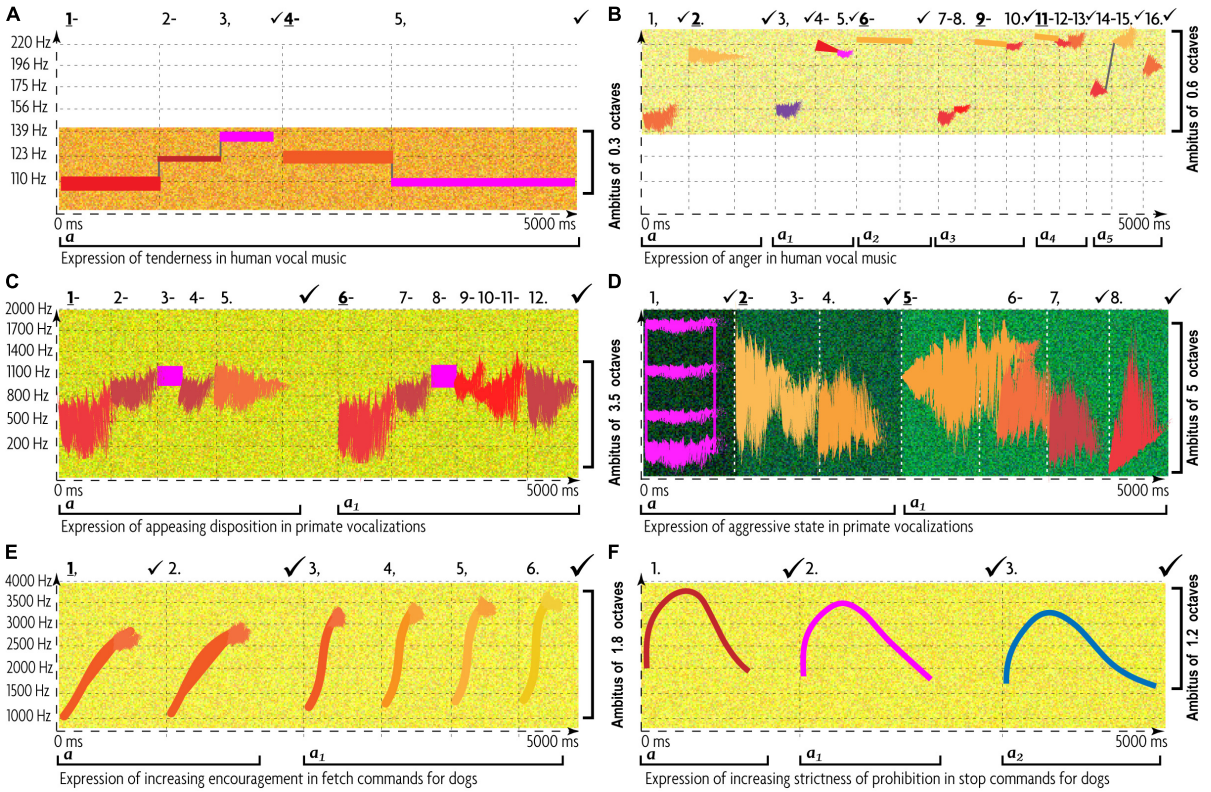



**FIGURE 5 |** Hybridization of characteristic patterns of ACs and human music in encouraging and prohibiting commands by human trainers to their dogs (McConnell and Baylis, 1985; McConnell, 1990, 1991, 2002; Miklosi, 2015). **(A) Typical expression of tenderness in human music**. This diagram extracts the key features of **Table 2** and **Figure 2**: very few pitch-classes with a low rate of change within a narrow ambitus, wave-like melodic contours filled by stepwise motion in the low register, slow tempo, with long tones and tendency to decelerate, and regular meter yet rhythmic diversity. Articulation is mostly legato, with occasional pauses. Dynamics is soft, stressing the anchor tones. **(B) Typical expression of anger in music** (according to **Table 2** and **Figure 3**): many pitch-classes with high rate of change and wide ambitus, ascending contours, and leaping zigzagging motion in high register. The tempo is fast, with short tones, often accelerating, with irregular pulse, and strong rhythmic contrasts. Dynamics is mostly loud, and accents fall on metrically weak tones. **(C) Typical expression of appeasing disposition in primate vocalizations** (**Table 3**). Many pitch-levels have a high rate of change, following a gradually ascending melodic contour within a relatively narrow ambitus. Tempo is fast, with short tones and long groupings. These features strongly contrast **(A)**, whereas metric regularity, legato articulation, low registration, and soft dynamics resemble **(A)**. **(D) Typical expression of aggressive disposition in primate vocalizations** (**Table 3** and **Figure 4**). There are relatively few pitch changes due to an extremely broad bandwidth, precluding frequent leaping. Long tones are embedded in fast motion with a descending contour in low register. These features oppose **(B)**, whereas meter, articulation, dynamics, and harmonicity resemble **(B)**. **(E) Typical expression of growing encouragement in fetch-whistles for dogs**. This expression combines a tender disposition of a human **(A)** with the appeasing disposition of a dog **(C)**. Therefore, fetch-command has to reconcile the contradictions between AEs' expressions of **(A)** and **(C)**. To accomplish this, the ascending contour becomes steeper, each signal and the time interval between signals become shorter, the ambitus of each signal grows and reaches higher register, and the groupings grow in size (from 2 to 4). Temporal and pitch AEs are co-adjusted, merging traits from **(A)** and **(C)**. **(F) Typical expression of growing prohibition in stop-whistles for dogs**. This expression combines the display of human displeasure, like **(B)**, with the appeasing disposition of the dog **(C)**, while structurally and semantically opposing **(E)**. **(F)** subverts a single long tone to the contrasting gradual flections in pitch, where the descending portion receives the greatest significance. The increase in intensity of prohibition is signified by extending the time values and reducing the steepness of the descending curve—in contrast to **(E)**. Dynamics provides yet another axis of opposition: loud for **(E)** versus soft for **(F)**. Most importantly, the **(E,F)** opposition involves a compensatory interaction of the temporal, dynamic, and pitch patterns of AEs. Thus, whenever **(F)** is used in isolation, its softness, slowness, and ametricity might project the impression of passiveness—contrary to the categorical nature of a "stop" command. To avoid this, **(F)**'s melodic curve combines ascending and descending curves whose conflicting relation generates extra tension.

## DOMESTICATION OF ANIMALS SETS THE NEED TO MAKE TONAL ORGANIZATION SEMIOTICALLY FUNCTIONAL

The need to command domestic animals underlaid the population explosion of both humans and livestock during the Neolithic Revolution. Animals benefited from human support, while humans benefited from animal produce. They both had to establish common patterns in their existing codes of vocal communication and adopt new patterns wherever the old patterns were deficient. Aspect-matching of pitch and rhythm was part of "bi-specific translation" of human commands (**Figure 5**). Rhythm reflects the "motion" pattern characteristic for a given "emotion" (Amaya et al., 1996), while pitch—the exertion/effort required by such motion—jointly defining a "sound gesture" (de Götzen, 2004). Perception of pitch and

rhythm relies on the biological components mutual for mammals, thereby supporting heterospecific communication. There is fMRI evidence of shared emotional vocalization systems across species (Belin et al., 2008).

An account of pitch-rhythm interaction comes from dog-training. Long continuous low/descending pitch is universally used to stop a dog, whereas repetitions of short rhythmic high tones—to encourage it—which might comprise a mammalian generality (McConnell and Baylis, 1985). Dog trainers identify pitch contour, rhythm, repetition rate, and amplitude as AEs effective in dog's commands.

Stop/fetch opposition reflects a multi-dimensional compensatory interaction of pitch, rhythm, and dynamics, mutual for both humans and canines. Some of the animal acoustic "universals" became appropriated into this bispecific communication, while others were overruled. Thus, across mammals, greater amplitude generally corresponds to a higher level of arousal (Briefer, 2012). However, it is only the fetch-command that follows this rule, whereas the stop-command, in contrary, adopts soft dynamics to subdue a dog (McConnell, 2002, 49–63). This overriding of the natural association between dominance and loudness highlights the fundamental difference between human and animal communications (Owren and Rendall, 2001):

- Human communication is "*receiver-centered*"—TO caters to information requirements of the listener;
- Animal communication is "*sender-centered*"—TO reflects the psycho-physiological state of the signaler, disregarding the listener.

Human-to-animal communication integrates both strategies:

- *Humans address animals*, treating them like humans, but *perfect the encoding* to secure the desired response. Thus, "doggerel" (Hirsh-Pasek and Treiman, 1982) constitutes dog-directed adaptation of human motherese (Mitchell R. W., 2001).

Pitch contour is a primary AE for most human cultures. Melody is the only aspect that differentiates between the basic musical emotions completely on its own (**Table 2**)[41]. In ACs, pitch does not provide such differentiation (**Table 3**). Pitch's importance for music pushes human melodies *higher* in register. This is because the low frequencies appear softer (Oxenham, 2013)—making the low contours less salient than the high contours. The same applies to primate hearing and, possibly, other mammals (Stebbins and Moody, 2011). Domestic animals too should follow suit. This incentivizes humans to raise contours characteristic for basic emotions above 1 kHz, where pitch changes are more salient. The only exception is the affection/love signals. Intimacy requires close-distance communication where the softness of low-frequency poses no problems.

Social animals share affective signaling system with humans (Snowdon et al., 2015). This enables effective musical communication between humans and domestic animals—all of whom are "social" (Stricklin, 2001). SFTO in all likelihood evolved gradually, following the schemata of human-to-dog communication. The earliest archeological evidence of domesticated dogs dates back to 15 kya (Larson et al., 2012), but signs of domestication were found in a Gravettian site, at Předmostìi (Germonpré et al., 2012). The DNA analysis indicates that a dog-like 33 kya old fossil from Altai is closer to modern dogs than to wolves (Druzhkova et al., 2013). Dog domestication must have been slow, preceded by feeding dogs with leftovers in exchange that they would follow humans and alert them of approaching predators. Dogs are genetically adapted to digest starch, which constituted part of human diet (Axelsson et al., 2013). Similar adaptation occurred in dog's communication system. *It adopted traits of human TO*. Compared to wolves, dogs use more vocal signals, especially bark-based—and barks feature co-modulation of two expressive aspects, amplitude and rhythm (Simpson, 1997). Alerting and territorial barking, both vary in intensity and rate depending on the distance of the dog from the conspecific or heterospecific intruder and the extent of the dog's arousal. At near distances barks become louder and more rapid. Such signaling and the manner of its modification most likely evolved in response to human's selective pressure on dogs to bark territorially at strangers (Simpson, 1997).

Human-to-dog communication most likely prototyped communication to later domesticates: cows, sheep, and goats. The surviving Nordic tradition of *kulning* provides the gist of the Neolithic pastoral music-making.

## THE SCANDINAVIAN TRADITION OF KULNING AS A MODEL OF NEOLITHIC MUSICAL SEMIOSIS

Animal husbandry in Scandinavia started ≈1800 BC and reached its "golden age" by 1200 BC. This is when owning larger stocks became prestigious while climate warming enabled outdoor animal maintenance almost year-long (Tesch, 1992). However, winter grazing was hard on bushes and trees, depleting local resources. This, along with subsequent climate cooling, brought about a new housing style, designed to shelter animals together with humans for winter—which characterized Scandinavian pastoralism (Armstrong Oma, 2013). Sharing the house with animals led to acceptance of animals as household members, equal to humans, and categorically as "clean"—even animal dung was used to make wattle and daub walls. Sharing is known to increase bonding. Human dependence on milk products, and animals'—on humans' "room and board" promoted mutual trust and attraction (Armstrong Oma, 2010). From being "products," animals turned into "producers" of dairy. This brought about *psychological "revolution"* in human-animal relationships, where music acquired the leading role.

[41] Prevalence of ascending contour characterizes happiness, anger, and fear. Happiness differs from anger and fear by employing variety of melodic contours called to diversify an ascending contour. Anger differs from fear by using sharp rather than wave-like contours and by dominance of staccato articulation in pitch changes (fear mixes staccato and legato articulations). Prevalence of descending contour characterizes both, sadness and love. They can be distinguished solely by intonation: flattened with stepwise falling contours for sadness, and sharpened with occasional ascending leaps for love.

Milking required concordance. An irritated animal or milkmaid reduced milk-yield, reducing human nutrition. Humans had to maintain mutual affection toward animals—evident in taboos on swearing/screaming at cattle, widespread across Eurasia (Plotnikova, 1999b). Music ritualized and fortified this union across different cultures (Shevtsov, 1988; Wallin, 1991; Alekseyev, 1995; Ivarsdotter, 1995, 2004; Novik, 1999; Dorina, 2004; Dissanayake, 2005; Kolltveit, 2008; Cheng, 2009; Yoon, 2018), especially evident in surviving traditions of milking songs (Nielsen, 1997; Pegg, 2001; Gioia, 2006b), animal lullabies (Kondratyeva, 1989; Kyrgys, 2002; Tchotchkina, 2003; Kan-ool, 2012), and spells (Kondratyeva, 1996; Kyrgys, 2002; Bordzhanova, 2007; Sodgerel, 2012, 2016; Tiukhteneva, 2017)—which all share the union of musicality and love/care that characterizes human motherese (Trevarthen, 2019).

Principal traits of such music can be extracted from the current practice of Scandinavian herder's music-making. Its chief task is to control the behavior of the grazing livestock during the warm seasons at distant pastures (Ivarsdotter, 2004). The herder aims at influencing the animal's emotional state over a range of distances, up to a few kilometers. Long-distance transmission requires a special vocal technique and musical instruments. The same musical signals convey different information to livestock and humans: commanding animals while informing animal-owners at the farmstead of their animal's wellbeing. This dual communication has been faceted through a transhumance system known as *shieling* in England (Cheape, 1996), and *fäbod* in Scandinavia (Svensson, 2015)—emerging during the late Bronze Age in response to the scarcity of local winter fodder (Tesch, 1992). In Sweden, the shieling standard was set in Dalarna, and the alternative local traditions are considered its variations (Svensson, 2015). Traces of shieling are spotted across Europe, from the Hebrides to the Carpathians, becoming widespread by the Iron Age (Cheape, 1996). In Norway, the earliest fossil fields of lynchets show signs of cultivation during the late Bronze Age (Skrede, 2005), confirmed by palaeobotanic and archeological dating (Kvamme, 1988).

Shieling is characterized by seasonal migration to a summer station where herders spend their daytime supervising animals, preparing fodder for the coming winter, and produce dairy during evenings (Cabouret, 1984). Since milking, butter- and cheese-making traditionally constituted the women's job, shieling and its music became female prerogatives in Scandinavia. There, milking could dishonor a man, and shieling was managed exclusively by young women (Svensson, 2015). In Ireland, shieling was a family business, whereas in Spain, France, and Switzerland dairy-work and herding were conducted by men.

The gender difference, undoubtfully, played a role in shaping the European pastoral musical traditions. Scandinavian, Icelandic, Alpine, Jurassic, Pyrenean, Apennine, Sardinian, Balkan, Turkish, and Caucasian mountains have sheltered singing styles that originated in the herding culture, and shared a peculiar singing technique based on a forceful high-laryngeal falsetto-like sound production (Wallin, 1991, 510). Wallin (pp. 511–23) summarizes the archeological, anthropometric, and genetic research to support the ethnographic findings of Carl-Allan Moberg (1971). Moberg outlines the core traits of the archaic *Fåbodväsendet* music: "head-voice" vocal technique, utilitarian function of long-distance signaling, and ideological roots in pagan magic.

The centerpiece of *Fåbodväsendet* tradition is its "maximal-distance" style—"*kula*"—that I distinguish from "kulning"—an umbrella-term for the entire *Fåbodväsendet*[42]. Local names for kulning (e.g., *lockrop*) imply the alluring of animals by magic properties of sound to suggest certain behavior to the herd, avert evil trolls and predator-animals—following shamanic tradition of maiden singing (Mitchell R. W., 2001). In Swedish mythology, forest spirits possessed their own cattle, and herdswomen (*kulerska*) learned kulning from *skogsrå*, "sirens of the woods" (Johnson, 1990). Suggestive power of kulning was deemed so high that women lived in *fåbods* alone without any weapons. Folk beliefs attributed this power to beauty. Indeed, well-ornamented high "warbling" register of distant female voice made men and women pause their work and enjoy the sounds (Ivarsdotter, 1986). For humans, kula clearly presented an aesthetic object despite bearing utilitarian status of "non-music" (Frödin, 1929)[43]. For animals, kula constituted a "safety call." Both attitudes focus on *positive* rather than negative emotions—not only to keep the cattle under human control, preventing panic, but also to boost the *kulerska's* confidence and alertness (Wallin, 1991, 420)[44]. SFTO must have emerged as a set of sonic attributes, perception of which was directly "wired" to reward circuits in brains of humans and domestic animals.

Wallin (1991, 420) rightfully stresses that matriarchy influenced early pastoralism: "the maternal instinct and care" instilled the social holding of attachment to stabilize and reinforce the animal-human affiliation. Distinctively female, *Fåbod* tradition must have prehistoric roots (Johnson, 1990). Motherese undoubtedly prototyped a close-range kulning. Animal-directed vocalizations acoustically and functionally resemble lullabies by commanding calmness/happiness—not just in Sweden (Wallin, 1991, 392) but also on the other side of Eurasia, in Altai (Kondratyeva, 1996). Common traits include

[42]There is a wealth of terms used in Scandinavian countries to refer to herding vocalizations (Rosenberg, 2003, 8). Although the term "kulning" (kolning) is most commonly used in English in relation to the special technique of the long-distance vocal calling, I follow Wallin (1991, 387) in reserving the term "kula" (he uses the alternative spelling "kola") which in Swedish means "to make a distant call" exclusively for long-distance communication. This is necessary, because long-distance "kula" calls are routinely inserted in mid-distance and close-distance vocalizations, while it is the long-distance "kula" style that distinguishes shieling vocalizations from other forms of traditional Scandinavian music.

[43]It should be noted that the peculiar status of pastoral music as a form of *heterospecific* communication is responsible for the emic views on kulning as **non-music**. This is yet another confirmation of the need in the etic approach. Across Eurasia, herder-made music is distinguished from "normal" music as a form of "magic." The profession of the herder is traditionally associated with sorcery: herders are believed to sign a contract with the evil forest spirits, receiving magic power for vocal and instrumental music-making in exchange for not using their gifts publicly, under the threat of death (Plotnikova, 1999b). At the eastern end of Eurasia, in Altai, supernatural beliefs are even stronger, reserved not only for professional herders (*chabans*) but for all livestock-owners who use pastoral spells (Kondratyeva, 1996). All vocalizations of this type are considered non-music—to the extent that informants perceive any request to "sing" a spell as being ridiculous.

[44]Noteworthy, despite a 16-hour-long workday and insecurity of living alone without any weapons, shieling jobs were always highly sought after, since women remained in charge of their summer life and enjoyed freedom unavailable to them at the farmstead (Rosenberg, 2014).

prolonged singing, formulaic regularity, vocables, smooth contours, motherese-talking, and caressing (Tiukhteneva, 2017). In animistic societies, both infant-lulling (Kondratyeva, 1989; Farber, 1990; Tchotchkina, 2003; Gioia, 2006a; Milne, 2017; Garroway, 2019) and domestication rites for newborn cattle (Aksyonov, 1964; Johnson, 1990; Kondratyeva, 1996; Plotnikova, 1999b; Kan-ool, 2012; Tiukhteneva, 2017) are associated with magic, achievable by female "charms."

Similar to lullabies are milking songs (Nielsen, 1997)—used across Eurasia, from Scotland to Mongolia (Gioia, 2006b, 71). Remarkably, when milking, Mongolian herdsmen switch to motherese-like "musical talk," based on animal onomatopoeia (Yoon, 2018). Known cases of male pastoral calling engage falsetto to imitate the female model (Uttman, 2002). Similarly, in surviving pastoral traditions of Altai, lulling is reserved for women, and require throat-singing if sung by men (Tiukhteneva, 2017). Pastoral spells in Altaic tradition constitute female prerogative, but are occasionally performed by men (Kondratyeva and Kopytov, 2017), engaging throat-singing (Kyrgys, 2002, 64). Like falsetto, throat-singing emphasizes harmonics that make melodies appear registrally higher—closer to the female range—and, like female kula, resembling pure tones.

The same applies to whistling signals, used across Eurasia by herdsmen to stimulate and/or safe-guard animals (Levin and Suzukei, 2006, 134–40). Just like kulning, in pastoral societies whistling is associated with sorcery (Plotnikova, 1999a) and is thoroughly regulated by taboos (Dzenzelevskii, 1984). Acoustically, whistling comes closest to "kula" in distance-range, loudness, and tonal quality (Eklund and Mcallister, 2015). To command their animals, Altaic herdsmen produce whistles audible over 4–5 km, and throat-singing—3 km (Pegg, 2001, 236). Curiously, female "head voice," required by kula, is called "whistle register" (Sundberg, 1987, 50). And *xöömii* (throat-singing) is considered a form of whistling in Mongolia (Pegg, 1992).

Wallin (1991, 523) sees shieling music as part of the prehistoric expansion of a novel herding culture northwest of Anatolia/Balkan/Caucasus toward Iceland, with its base in Jamtland (**Figure 6**). Jamtland's "forest barrow" marked the end of tundra after the glaciers' retreat, attracting hunters and supporting a mixed pastoral economy that survived at the coldest outskirt of Europe practically unchanged until the late Middle Ages. Geographic and chronological distribution of cattle-herding across Europe, quite well-studied, provides timing references for Wallin's model. The outcome of this geomusicological[45] correlation is presented in **Figure 6**.

Domesticated cattle spread East-to-West along the Mediterranean coastline, encapsulating most of "yodeling" territories ≈6000 BC. The South-to-North expansion took much longer—Central Sweden became pastoralized in the 2nd millennium BC. Dissemination of cattle and Indo-European languages went hand by hand. The Indo-European language family covers most of Europe—except for Finno-Ugric languages of Fennoscandia and Russia. Another notable exception is Turkey whose Indo-European languages (Hittite, Luwian, Palaic, Lydian) died out during Antiquity. Formation of each new Indo-European language seems to have followed the adoption of husbandry. The yodeling areas correspond to the earlier stages in expansion of the Indo-European languages, conserved by the mountain systems: Taurus, Pontic, and Armenian Highland in Turkey, the neighboring Caucasus, Balkan, and more remote Carpathian, Alps, Jura, Apennine, Sardinian, Corsican, and Pyrenean. The dissemination routes either curve around the mountains or cross them by riverbeds. The oldest routs ran by the Mediterranean coastline along the 40N latitude, supporting the conclusion of Diamond and Bellwood (2003) that the domesticates and languages spread faster to East-West than to South-North. This explains the divergence of pastoral music tradition into two types: Southern yodeling versus Nordic kulning and kulning-likes[46], distinguished by different bovine genomes. Studies of Y-chromosomal variation have identified two primary taurine haplogroups in Europe, split in two homogenous regions alongside cultural, historic, religious, and linguistic boundaries between the pied or red cows of the Nordic and Baltic/Slavic lands, on the one hand, and the spotted yellow or brown breeds of Switzerland and southern territories, on the other hand (Edwards et al., 2011).

Kulning and yodel form respectively Northern and Southern "dialects" of a cattle-directed "language"—a satellite of the proto-Indo-European. The main role in the Indo-European "domestication package" belonged to cattle—the largest meat- and milk-source of all domesticates. The emergence of cattle-related mythology reflects the importance of cattle and explains the sudden proliferation of cattle burials across Northern Europe ≈3000 BC (Sjögren and Price, 2013)[47]. Symbolic elevation of cattle could characterize the entire Neolithic "revolution" in Eurasia, more noticeable in Scandinavia, where ox symbolism replaced red-deer symbolism after ox overtook deer as the most important food source (Tilley, 1996, 183–4). If wild deer opposed the human sphere as a *utilitarian* object of desire, domesticated ox was included into the human sphere as the *emotional* object of desire. And music is indispensable in supporting emotionality.

Divinization of music (Franklin, 2006) and ox (Campbell, 2017), so prominent in Indo-European tradition, could have a single origin in Indo-Iranian lands—bound to the concept of non-violence (Tull, 1996). Cattle sacrifice is depicted in prehistoric Sujanpura petroglyphs (Brooks and Wakankar, 1976). The ritual use of burnt cow dung is still common in Hinduism,

[45]The scope and the method of geomusicology were introduced by George Carney (Nash and Carney, 1996). Izaly Zemtsovsky formulated an analogous approach in his proposal to establish a new discipline of ethnogeomusicology (Zemtsovsky, 2005).

[46]Thus, Finnish "ringing calls" present a form of vocalization that acoustically and culturally resembles Swedish kulning while featuring a few unique traits (Uttman, 2002). Occasionally, ringing calls are performed by men (falsetto), utilize a peculiar lip technique (generating the "phui"-like tonal quality), and engage "darker" vowels.

[47]Cattle definitely carried special symbolic significance in Neolithic England (Ray and Thomas, 2003). Cattle received the same funeral treatment as humans in Danube winter burials as part of the Sun cult which thrived throughout the 4th millennium BC, probably because of drastic swings in solar activity (Horvaith, 2012). The second millennium BC Linear-B tablets from Knossos testify that, unlike sheep/goats, cattle was given names, bestowed with individuality—and was associated with royalty and sacrificial rites (McInerney, 2010, 50–53).

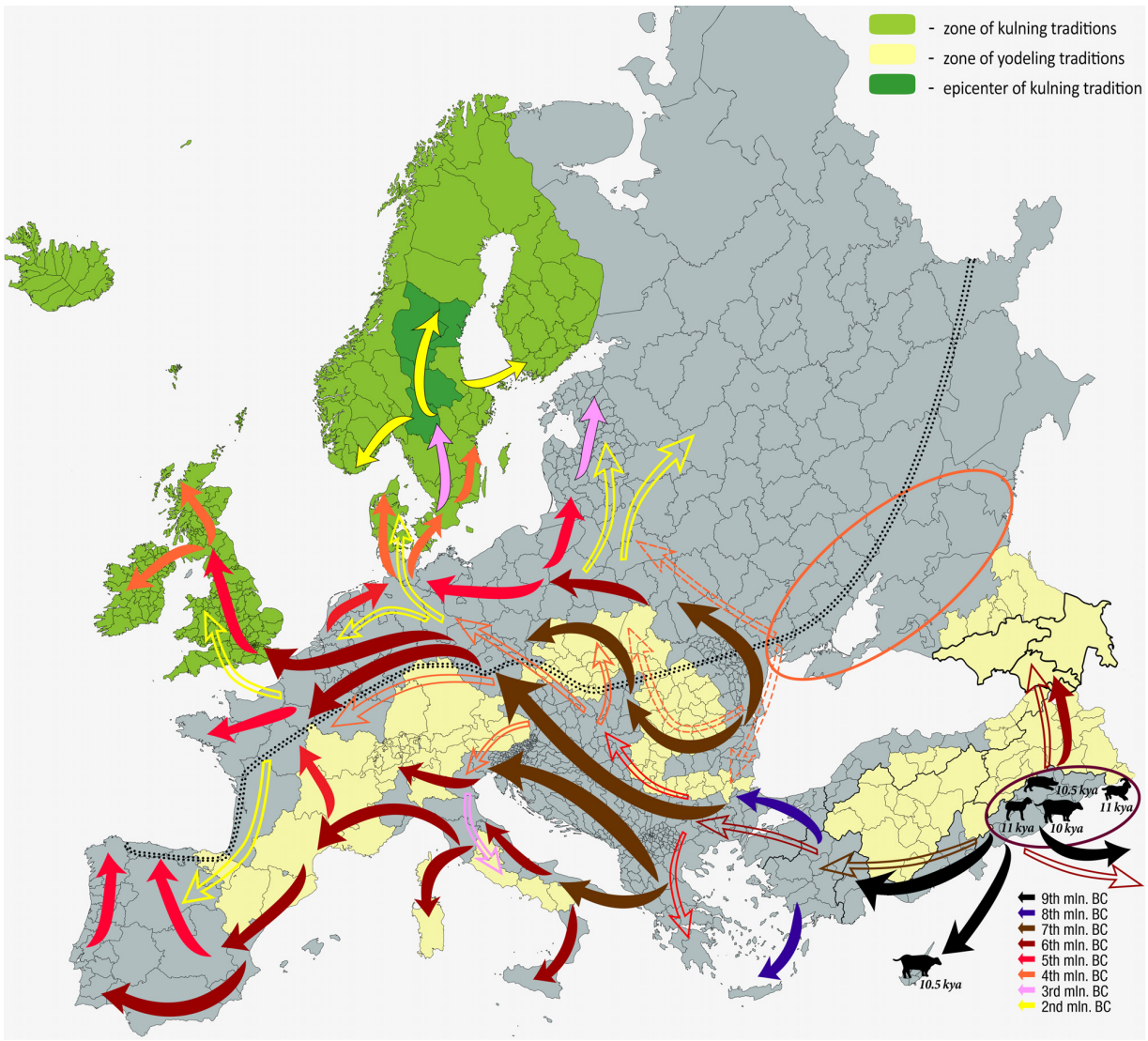



**FIGURE 6 |** The earliest spread of pastoralism across Western Eurasia. This figure shows the approximate timeline and the geographic correspondences between locations of herding falsetto-like vocalization, the oldest areas of cattle-breeding and distribution of Indo-European languages. Light green color marks the territory of shieling pastoralism, dark green—the "core" Fåbod regions, and crème—the area where yodel-like vocalizations survived within pastoral cultures (Moberg, 1955, 1971; Baumann, 1976; Leuthold, 1981; Ivarsdotter, 1986; Wallin, 1991; Mitchell S. A., 2001; Uttman, 2002; Plantenga, 2004). The origin of the latter can be dated by the timeline of the spread of domesticates over Europe, which is well studied. Animal icons show the approximate place and time of origin of domesticated cow, goat, sheep, and pig, based on available archeological data (Zeder, 2008; Driscoll et al., 2009; Peters et al., 2017). Color-filled thick arrows show the timeline and main routs of dissemination of domesticated cattle during the Neolithic and early Bronze Age according to the archeological and genetic data (Caramelli, 2006; Lõugas et al., 2007; Zeder, 2008; Rowley-Conwy, 2011, 2013; Tresset and Vigne, 2011; Bläuer and Kantanen, 2013; Marciniak, 2013; Saña, 2013; Schulting, 2013; Sjögren and Price, 2013; Berthon, 2014; Cramp et al., 2014; Felius et al., 2014; Sørensen and Karg, 2014). The darker the arrow's color, the older the date. The double-dotted black line approximates the border between the Northern and Southern European bovine genetic funds. Colored ovals and outlined arrows indicate the hypothetical origin and the spread of Indo-European languages according to the computational methods, based on Bayesian logic and phylogenetic analysis algorithms (Diamond and Bellwood, 2003; Gray and Atkinson, 2003; Atkinson et al., 2005; Atkinson and Gray, 2006; Bellwood, 2008; Gray et al., 2011; Anthony and Ringe, 2015; Chang et al., 2015; Heggarty, 2015). The brown oval marks the area of genesis of Proto-Indo-European language according to the "Anatolian hypothesis" (Renfrew, 1987), whereas the orange oval—to the earlier "steppe hypothesis" (Gimbutas, 1993; Anthony, 1995). The dashed outlined arrows show the earliest stages of dissemination of the Indo-European languages from the Yamnaya epicenter. Both hypotheses generally agree in defining the later stages (Gray et al., 2011)—represented by solid outlined arrows.

traceable to the 3000 BC Ashmounds (Boivin, 2004). The Shiva-bull affiliation is evident in the Bronze Age Harappan "Proto-Shiva" (Hiltebeitel, 2011). Harappan symbolism clearly elevates the cattle over other domesticates, evident in the buffalo figurine amulets and seals that are likely to assimilate the west-bound Indo-Iranian cult of Mother Goddess, eventually forming the "Sacred Cow" concept (Lodrick, 2005). This corresponds to veal and cow-milk becoming primary foods during Rigvedic and

Vedic times—there were people at that time who lived on milk alone (Prakash, 1961, 12). Milk products were used in rituals and offerings to gods, certainly accompanied by music, promoting the transformation of cow into the symbol of femininity and fecundity in Vedic literature (Brown, 1964). Consecration of cow gave it purity: even its urine and dung were used for healing and cleansing (Korom, 2000).

The cultural context of kulning and the tradition of home-sharing with cattle strongly resembles the Vedic cultural blend of non-violent femininity, cow-worship, and magic. It is not accidental that kula finds a nearly perfect match in Tibetan traditional pastoral songs with long rhythmically free undulating phrases, extremely tense timbre of high quasi-falsetto voice, generous ornamentation, and an ongoing variation (Stuart, 2008, XXIV). This is the most ancient of the three major forms of Tibetan music, peculiar to a nomadic pastoral culture, and originating from cattle calls (Crossley-Holland, 1967). Like kulning, it incorporates parlando and recitative for close-distance vocalization to animals, and also includes milking songs (Plantenga, 2004, 113).

Introduction of milk revolutionized the Neolithic lifestyle, supporting the psychological revolution in human-animal relations and bi-specific musical communication—especially in Northern Europe, where milk quickly replaced fish as the main food—manifested by the widespread adoption of milk-storing pottery (Cramp et al., 2014). The archeological evidence agrees with the genetic evidence of the time of emergence of lactase persistence[48]. Lactase persistence reflects the adaptation to diet (Hancock et al., 2010)—without which adults have lactose intolerance and nutritional loss (Campbell et al., 2005). Ill effects of malnutrition coexisted with milk-bound diseases during the adoption of the milk-based diet. Mycobacterium tuberculosis existed 40,000 years ago, but became pathological for humans only from 6200–5500 BC onward (Hershkovitz et al., 2015) - by the time when the spread of husbandry reached Central Europe. Seemingly "the same" milk could either kill or nurture life—which must have promoted new supernatural beliefs and rituals to "exorcize" milk-production in replacement of the earlier hunter/gatherer rituals. Music, so common for religious applications, most certainly supported this reform.

For Europe, geographic distribution of Indo-European languages[49] (Heggarty, 2015) goes hand in hand with the distribution of taurine mtDNA that descends from the Fertile Crescent (Caramelli, 2006). And subdivision of the bovine European genetic pool into Northern/Southern genotypes (Edwards et al., 2011) matches the distribution pattern of lactase persistence: 40% of adults in Greece versus 90% in Scandinavia/England (Curry, 2013). Those populations that consumed more dairy have higher occurrence of lactase persistence (Bersaglieri et al., 2004). Evidently, milk dependence was more than twice higher in the North. The Indo-European expansion occurred through the farmers' immigration and interaction with local foragers rather than by technological import alone (Rowley-Conwy, 2011). Greater lactase persistence in the North reflects the dairy's effectiveness in providing nutrients, the convenience of its storage in cold climate, the insurance against bad harvests (Gerbault et al., 2013), and health benefits of increased vitamin D consumption in low-sunlight conditions (Flatz and Rotthauwe, 1973).

Kulning emerged to nourish the symbiotic co-dependence of humans and cattle in harsh Nordic conditions that demanded stronger bonding than those of more diverse pastoral economies of Southern yodel territories, therefore employing a **female pastoral model**.

The biggest contender for the Indo-European language family in Northern Europe—the Uralic family (Diamond and Bellwood, 2003)—relates to another domesticate: the reindeer. Reindeer hunting was essential for colonization of Eurasian Arctic/Subarctic (Gordon, 2003). However, reindeer domestication still remains in its early phase (Reimers and Colman, 2009). The distinction between reindeer-hunting and reindeer-herding remains vague—even reindeer owners often do not know if a particular reindeer is "wild" or "domestic" (Ventsel, 2006)[50]. Leading fences and corrals have been used for hunting wild reindeers and only recently have they become "domestic" accessories (Aronsson, 1991). Reindeer pastoralism emerged gradually from taming individual reindeers for transportation and decoy-hunting—compensating for the depletion of wild reindeer population (Vorren, 1973) that occurred during the 13–16th centuries (Hansen and Olsen, 2014, 175)[51]. Reindeer domestication must have started in parallel with cattle domestication in Norway/Sweden but lingered into the Middle Ages—absorbing cultural traits of human-to-cattle communication.

The principal psychological trait of kulning is the "humanization" and child-like patronizing of cattle. Similar attitude characterizes reindeer pastoralism: animal is treated like a family member whose life is valued and its attitudes are respected (Ingold, 1986). Kulning, yodel, and reindeer-communication should all be regarded as various "**languages of domestication**," generated by borrowing "acoustic traps and snares"—i.e., onomatopoeic decoy calls—from hunters

[48]Lactase persistence was completely absent in early Neolithic population 5500 BC (Burger et al., 2007), making its first appearance in Scandinavia in 3400 BC (Malmström et al., 2010), by 3000 BC in Iberia (Plantinga et al., 2012) and taking over Europe thereafter (Marciniak and Perry, 2017). This timeframe agrees with the scenario represented in **Figure 6**.

[49]The Indo-European family contains 144 languages divided amongst 11 distinct branches—with even more languages most certainly having existed in the past but gone extinct (Diamond and Bellwood, 2003). In Europe, non-Indo-European languages are limited to merely 11 documented languages (only 8% of the total number of languages): Etruscan, Basque, Iberian, Tartessian, Estonian, Finnish, Urartian, Sumerian, Hurrian, Hattic, and Mitannian—plus 3 undocumented languages: Pictish, Lepontic, and Ligurian (Robb, 1993).

[50]Herders routinely let their reindeers graze unsupervised for a rather extensive length of time. Inevitably, many animals become lost, turn wild, and can then be hunted (Stépanoff et al., 2017). Also, the herder's strategy of searching for his lost animals strikingly resembles that of hunting.

[51]This caused the import of non-native reindeers via the emerging Russo-Finno-Scandinavian markets and transition to pastoralism (Røed et al., 2018). Genetic evidence points to 3 epicenters of reindeer domestication: Fennoscandia, Western and Eastern Russia (Røed et al., 2008). Reindeer domestication took about 6000 years. Its earliest evidence comes from the 4000 BC petroglyphs (Helskog, 2012), a 1510–1130 BC burial (Murashkin et al., 2016), and the paleolinguistic tracking of words for reindeer that date back to 1500–1000 BC (Aikio, 2006).

and syntactically reorganizing them into "animal-directed" words to control the herd, its leader, and the individual animals (Alekseyev, 1995).

Kulning and yodel are *Indo-European* musical "*cow-languages*," later adapted for goats/sheep as they became personalized like cows[52], whereas reindeer-vocalizations make a *Finno-Ugric* "*reindeer-language*."

Kulning's SFTO was forged by long-distance delivery of the desired subharmonic structure. Kula is characterized by dynamic maximization (80–100 dB SPL at 50 cm)[53] while fixing 4 formants at FF, 1700, 3,000 and 4,000 Hz throughout all frequency changes, restraining vibrato, and raising the larynx above the resting position (Johnson et al., 1982). Elevating laryngeal position up to 4 cm increases the sub-glottal pressure tenfold as compared to talking (Ivarsdotter, 1986). Somehow, this causes no distortions, and kula's "harmonic signature" remains virtually unchanged at close- and mid-distances (1–11 m)—contrasting the "classic" falsetto (Eklund and Mcallister, 2015). Harmonic conservation is still observable at 22 m in kulning, albeit varying between different performers (Eklund et al., 2019). Evidently, kula is designed to transmit *kulerska*'s harmonic and melodic "signatures" to the herd at distances common in herding (Rosenberg, 2014).

Long-distance spectral optimization is known in intergroup communication of some primates (Waser and Waser, 1977). However, optimization to preserve subharmonic structures is unique to kula.

Kula's sounds are supposed to stand out in the environmental soundstage by featuring unnaturally hyper-periodic noise-free spectrum. Kula's harmonicity aligns with "**pleasantness**"—following the cow-bell paradigm. Animal-bells were used in Scandinavia at least from 1–4th centuries (possibly, from the beginning of the Bronze Age) to repel evil spirits, mark a human-controlled territory, and decorate the herd's leading animal (Kolltveit, 2008). For cattle, the bell signified human control, herd-leader's authority, and a safety signal. Humans associated bells with nature, peacefulness, goodness, and protection, employing bells to "borrow" the land from the forest spirits (Emsheimer, 1991, 43). Therefore, overall harmonicity signifies strongly positive values—in line with kulning's perceived beauty and safety/care. Across the animal world, too, harmonicity (pure-tonedness) and inharmonicity are meaningful along the friendliness/fear opposition (Morton, 1977).

Long-distance transmission requires high intensity and register. For 1 km, the most effective transmission occurs at ≈2 kHz (= C7) (Graf, 1980; Gray and Atkinson, 2003)—the range of a piccolo flute. Perhaps, whistling prototyped kula. Whistles are common in communication with dogs and the herd. And whistles exceed calling and yodeling in long-distance intelligibility: correct identification of whistles at 170 m distance is 95% versus 58% for yodeling and 70% for calling (Titze et al., 2018). *Bi-factorial* changes of rhythm/pitch-contour in whistling signals would pave the road for *tri-factorial* changes of rhythm/pitch/phrase-length in kula.

Long-distance communication eliminates mimics and gestures from semiosis, making it rely exclusively on acoustic attributes and demanding long-term memory (Wallin, 1991, 390). Exclusion of visual cues promotes the prolongation of a musical expression to facilitate its recognition and memorization. Therefore, phrase length reflects the distance: longer distances require longer phrases (p. 391). Changes in distance generate musical syntax (**Figure 7**).

*Close distance* promotes short phrases of multi-registral motherese-like recitative where only the "reciting tones" are pitched, and exaggerated leaps employ legato and portamento (**Figure 7B**). Pitches have tendency to monotony in low register at phrasal ends, which generates tonicity. Vocalizations are mostly stimulating and diverse in their referential/propositional content.

*Middle distance* makes motherese inaudible, instead requiring a different approach. Vocalizations become euphonized: engaging "parlando" rather than recitative[54], "smoothening" the leaps, increasing the share of pitched tones, and stressing rhythmic patterning and ordering. The calming effect of these adjustments, inappropriate for stimulating applications that are mostly common for mid-distance communication, is compensated by intensifying dynamics, structural contrasts, and staccato articulation (**Figure 7A**). Notwithstanding diversification, the highest-register "peak-tones" at motivic beginnings are often monotonous, prototyping the musical "leading-tone" by requiring some sort of continuation (as in a melodic resolution).

*Longer distance* further increases the share of musicality and pleasantness in herding vocalizations. They prioritize audition over visualization by engaging "call-phrases," made of exclamatory imperatives and summoning, free from referential/propositional context (Wallin, 1991, 417). Verbalized vocalization is replaced by a wordless kula (p. 410). Simple phrase-sentences consist of motif chains akin to incipits, climaxes, and cadences of Gregorian tunes (Helmer,

[52]Ivarsdotter describes how goat-calling follows the model of cow-calling, adapting it to the livelier nature of goats, notorious for their proneness to naughtiness (Ivarsdotter, 2004). Similarity of cow-calling (*Kolock*), goat-calling (*Getlock*), and sheep-calling (*Fårlock*) is obvious from listening to their archive recordings published by Swedish radio (Ivarsdotter, 1995). The same similarity is retained in pastoral incantations and spells that survive in Altai region—all three types of calling differ primarily in the prevalence of different phonemes for each of these three animals (Tiukhteneva, 2017). The musical characteristics of all three types of calling closely resemble one another (Kondratyeva and Mazepus, 1999). This suggests that similarity between cow, goat, and sheep pastoral communication is a wide-spread Eurasian phenomenon.

[53]The highest SPL level is reached at a 30 cm distance from the sound source (125 dB) which exceeds the ear's pain threshold at 120 dB (Rosenberg, 2014). The average SPL of kula at 1000 Hz is 113 dB. This is dynamically comparable to an operatic soprano singing fortissimo, except that the soprano's technique requires maintaining a fixed larynx configuration at a *low* position. However, the maximal SPL of the soprano does not exceed 90 dB near the lips and does not change much from modulating the pitch (Johnson, 1984).

[54]The term "parlando" was adopted by Anna Johnson in her report (Johnson, 1979) despite the traditional use of this term to refer to Western operatic singing that imitates speech and engages speaking "voice registers" (Sicoli, 2015) despite the absence of such intention for *kulerska*. Sung out words of close-distance kulning surprisingly resemble the operatic "parlando" sound. The kulning parlando contrasts the recitative kulning that minimizes voicing and remains much closer to talking than to singing, especially in its dynamics. The opposition of kulning parlando and kulning recitative resembles the opposition of operatic parlando and secco recitative, on the one hand, and the genre of melodrama that became popular in Western classical music in the 19th century, on the other hand.

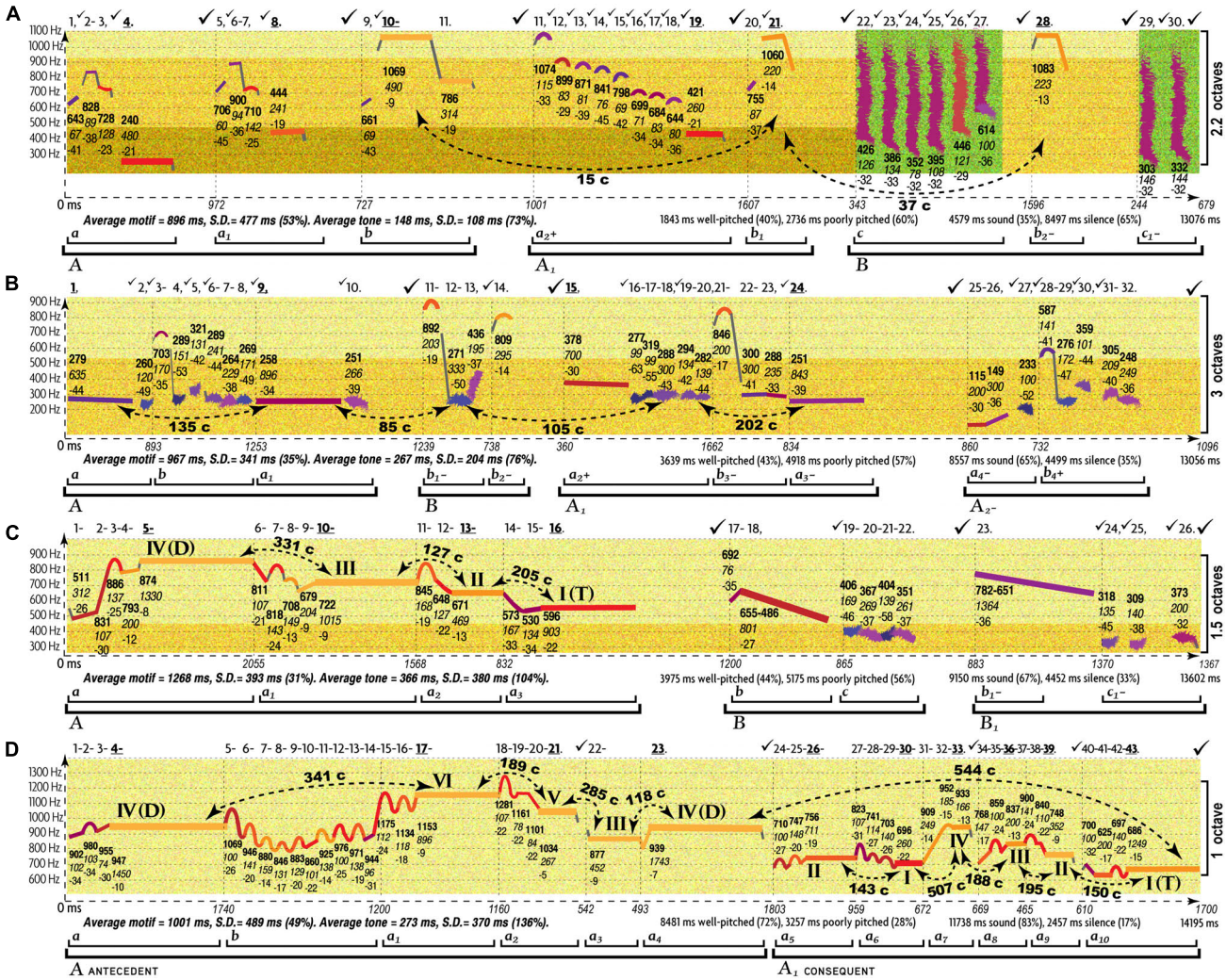



**FIGURE 7 |** Patterns of TO in four main types of vocalization in the vocal tradition of kulning. Since kulning is essentially ametric and averbal (except for the closest range recitative), its analytic charts omit lyrics. Unlike the previous figures, the vertical dash lines indicate the onset of motifs. The colored arc-line symbol represents an ornamental melismatic shake. **(A) Stimulative medium-distance kulning: parlando (a), exclamation (b), and onomatopoeia (c) motifs** (http://chirb.it/ntlxfM). This style is designed to compel the entire herd to move in the desired direction and, most probably, sets a model of interaction with animals for the other three styles. The three motifs achieve stimulation, each in a different way, contrasting one another in register, harmonicity, rhythm, and articulation. Motif "a" alerts by its staccato zigzag leaping between two registers. Motif "b" combines stimulation (staccato leap up to the "shrieking" register) with relaxation (legato leap down to the long tone). The "shrieking" peak-tones maintain the same pitch level (melodic regularity)—reflected by the dotted double-arrows (numbers indicate the frequency discrepancy in cents). Motif "c" teases the cattle by imitating dog's barking. The stimulative specialization of **(A)** is manifested in its prevalence of staccato, loud dynamics, three registers within a wide ambitus, exuberance of leaps, and briefness of motifs and tones. Noteworthy, the motifs "$a_2$" and "c" resemble the "fetch-command" archetype (**Figure 5E**). **(B) Stimulative close-distance kulning: recitative (a) and motherese (b) motifs** (http://chirb.it/8K3Lqg). **(B)**, like **(A)**, is stimulative but dynamically gentler due to closer distance (≈9 dB softer). This allows for diverse motherese-like prosodic exaggerations in motivating individual animals. Motif "a" expresses love/care by greatly prolonging the "recitative tone," sustaining its pitch and harmonicity. Motif "b" stimulates animals by briefly stressing the upper "head-voice" register with a shake-like embellishment, then sliding it all the way down to the low talking voice. Compared to **(A)**, **(B)** is smoother: fewer registers, less staccato, and longer motifs and tones. **(B)** tends to support a monotone (a predecessor of tonicity), most noticeable at phrasal ends. **(C) Inhibitive longer-distance kulning: simple kula (a), exclamation (b), and parlando (c) motifs** (http://chirb.it/n6f0sv). This style functionally opposes **(B)** by commanding the herd to stop grazing and to go home, implying that it is no longer safe to stay out. The chief function of "a" (kula) is to instill confidence in the herder's control over the animals. "Kula" typically consists of a chain of motifs stitched together to form a characteristic shape of steep ascension to the crest point and thereafter a gradual fall-off. However, motifs might differ according to their phrasal functions: initiation, climax, decay, and cadence. The resulting kula receives a basic modal TO: anchor tones constitute "degrees" of the mode, forming a fifth between the marginal degrees and dividing it in wider upper and narrower lower parts. The Roman numerals indicate degrees (I = stable is marked as T = "tonic"). The "b" motif presents "exclamation": a gradual sliding down (≈4th), softer than in **(A)**, and shaped like the "stop-command" (**Figure 5E**). Similarly shaped is the parlando "c" motif, much smoother than **(B)** due to its prevailing legato, freer rhythm, more homogenous registers, and longer motifs and tones. **(D) Tropotrophic maximal-distance kulning: exclusive use of complex kula sentences** (http://chirb.it/gpyC7t). Delivering signals over a kilometer requires taking multiple short caesuras throughout the span of the kula's descending formula, which distinguishes **(D)** from **(C)** by making kula complex. Motifs make up phrases, and phrases—sentences, all of which create modal complexity: anchor-tones form intervallic relations that define degrees within a mode (usually, 5–7 degrees). Upper degrees open kula, forming an antecedent cadence (marked by letter "D"—"dominant" function). Lower degrees end kula with the consequent cadence (marked by "T"—"tonic"), providing resolution.

*(Continued)*

**FIGURE 7 |** Continued
Compared to **(C)**, sentences in **(D)** are longer, rhythmically freer, more homogenous (by maintaining legato, a single register, the narrowest of ambitus for all kulning styles, and no leaps). Relaxation, secured by modal resolution, is supported by beautification: exclusive use of legato in smoothly shaped phrases and exquisite ornamentation (shakes, trills). **(D)** differs from **(C)** by sacrificing dynamic shaping on a phrasal level and, instead, reproduces the same dynamic contour on a motivic level—the final long tone is almost always the loudest in a motif (i.e., stable). Increased homogeneity and melodic consonance (i.e., absence of leaps) are called to motivate the herd not to depart any further beyond the range of hearing kula.

1975). Each phrase is distinguished by a wavelike melodic-dynamic "envelop" with an abrupt quick rise and a gradual prolonged fall. Kula pushes vocalizations higher, squeezing their ambitus, homogenizing timbres and legato articulation, while loosening the rhythm (**Figure 7C**). This triggers the modal genesis: kula's anchor-tones turn into degrees, with more-or-less sustained pitch values. The lowest degree becomes "tonic," in contrast to the unstable upper degrees, thereby forming tetrachord-based modes.

*Maximum-range* communication complicates kula by introducing hierarchic structuring (motifs-phrases-sentences) and by engaging the contrasting phrasal functions (initiation/climax/interruption/termination). The stimulating effect of the increased syntactic contrasts, undesirable for maximum-range communication that focuses on keeping the animals calm, is compensated by greater melodic homogeneity: maximizing legato, sentence-length, and dynamics, while minimizing melodic-intervallic, rhythmic, and registral diversity (**Figure 7D**). Longer span necessitates inter-phrasal caesuras, marking multiple phrases within long sentences, joined by stereotypical declining inter-phrasal melodic and dynamic "envelops." Melody relies on pentachordal skeleton, divided in upper major and lower minor 3rds, often supported with quartal/quintal infrafix (Johnson, 1979). Kula breaks in a series of antecedent-consequent sentences that engage different pentachord/tetrachord(s)—usually conjunct. This produces heptatonic modes (**Figure 8**).

The ongoing unveiling of musical structures makes kulning particulate by stacking up certain phrasal types while avoiding certain other combinations. This establishes syntactic rules and implicit music theory of TO for herders and herds. Herders perceive kulning as improvised "musical work in progress" (akin jazz improvisation) that elaborates a specific "theme" selected by the *kulerska* (Rosenberg, 2014). Herded animals probably perceive kulning as a series of programmed Pavlovian-conditioned routines. In both cases, compositionality promotes particulate semiosis: the meaning of a streak of phrases consists of the sum of the meanings of each of the constituent phrases. In effect, kulning tells a "continuing story" of the day, going through an elaboration of a musical theme (Rosenberg, 2014).

The herd's daily movement generates SFTO by stitching/restitching phrases of 5 syntactic-semantic types (**Table 6**):

- Kula (tropotrophic),
- Exclamations (phatic),
- Onomatopoeia (ludic),
- Parlando (imperative),
- Motherese/recitative (endearing/motivating).

Genesis of SFTO follows the path of human-to-dog whistling communication. Noteworthy, kulning's exclamations and onomatopoeic calls engage stop- and fetch-whistle features (see **Figures 5E,F**).

The proof for SFTO's pragmatic efficacy is in the herd's fulfilling of the shepherd's commands (Wallin, 1991, 410).

Yet another source of semiosis for kulning was phonemic symbolism. Complete absence of words in kula and minimal wording of motherese suggest the *prelingual* existence of kulning. Wallin (1991, 410–413) rightfully emphasizes that there is no reason to label kula's sounds as "*phonemes*"—they are mere *homologues* to vowels and consonants, shaped by the anatomic-physiological conditions of breathing and acting while uttering. The same applies to traditional Alpine yodel (Fenk-Oczlon and Fenk, 2009a). Yodel and kulning vocables are formed not by phonological oppositions of local languages but by the communication distance and the extent of the desired stimulation/inhibition for a given call. Thus, the highest larynx and intensity at the onset of long-distance inhibitive kula-phrases generate a semantically "negative" [i], whereas a relaxed post-climactic position in the mid-distance tropotrophic kula generates a "positive" [å]. Similarly, glottal stops at phrasal beginnings and endings range from a gentle [h] to a harsh [tj], depending on the needed attack and tenuto decay (Rosenberg, 2014). The choice of the most common kulning syllables (Ahlbäck, 2007) can be explained by human/animal's natural selection for effective distant communication (Wallin, 1991, 390).

**Monodization** of kulning was imperative in genesis of SFTO.

Animal communication usually employs male "chorus," male-female "duetting," or "antiphonal" formats (Yoshida and Okanoya, 2005). Musicologically, this corresponds to a special type of texture—"isophony": the ongoing out-of-sync multi-part reproduction of the same thematic material (Nikolsky, 2018). Isophonic jumble precludes SFTO. For multifactorial patterning to emerge, each vocalizer must clearly hear his/her voice in order to track spectral changes without any contamination by a partner. This is how infants learn to make their own songs and how children acquire "musical ear" (Nikolsky, 2020). Even in non-European traditions that are exclusively *polyphonic*, such as Aka Pygmy, motherese and children-made music remain *monophonic* (Rouget, 2011). This is because an auditory stimulus must be objectified to become accessible for reproduction: a relation of 2 tones in certain aspect must be realized as an auditory constant to lay the foundation for construction of a musical mode (Nazaikinsky, 1973). For perception, the listener must discover permanence of the foreground "sound-object" against the background of a sound-stage, and memorize it in order to relate to it all of the subsequent changes in the thematic material.

Just as one cannot learn prosody of a language by listening to the crowd, one cannot learn SFTO by listening to isophony. And

**TABLE 6 |** Acoustic traits of main motif types and their semantic values in kulning.

| Acoustic domains | Aspects of music | Kula singing phrases and sentences | Exclamation calls | Onomatopoeic imitations | Parlando singing and calling | Motherese recitative |
|---|---|---|---|---|---|---|
| 1. Frequency | 1. Melodic pitch (consecutive) | Prevalence of long descending stepwise motion; narrow ambitus of a 5th breaks in two portions: upper major. 3rd and lower minor. 3rd (with possible infrafix). Well-defined pitches, many melismas. The strongest melodic coherence. "Safety call" function –calming yet keeping alert, under control. Subject to aesthetic evaluation. | Prevalence of a zigzag pitch contour in a brief group of tones with a very short start, poorly defined pitches, medium long leaps, large ambitus of 0.5–1.6 octave. Optional but frequently employed shakes and melismas distinguish the stimulation use from inhibition, both engaged exclusively in the phatic function (no aesthetic value). | Short repetitions of the same imitation of an animal call: poorly pitched, very broad bandwidth, no melismas, melodic coherence is absent, stimulating function (to make the herd move in the necessary direction). Most likely, this is a derivative of a "fetch whistle" command. Utilitarian application (no aesthetic value). | Prevalence of a zigzag shape, like exclamation, but longer and stressing a huge descending leap (up to 1.6 octave), greater than an ascending leap (<7th). Both leaps relax toward the motif's end. They stimulate or inhibit, based on the extent of such relaxation. Utilitarian application with aesthetic value. | Prevalence of drastic contrasts in pitch between flat monotonous pitched talking and extreme zigzag leaping (about 1.7 octave) that is usually embellished with melismas. Stimulating and motivating functions, supported with gestures and mimics. Only a few pitches are clear. Mostly utilitarian application. |
| | 2. Harmony | n/a | n/a | n/a | n/a | n/a |
| | 3. Form (complexity) | Greatest complexity | Simplicity | Greatest simplicity | Low complexity | Medium complexity |
| 2. Time | 4. Tempo | Slow, frequent ritenuto toward the phrasal end. | Moderate and/or slow tempos, possible ritenuto. | Moderate, possible accelerando. | Moderate or slow, frequent rubato. | Moderate and lively, with moderate rubato. |
| | 5. Rhythm | Sharp contrasts of melismatic and anchor tones, the greatest rhythmic diversity, no increments. | Contrast of short-long pattern in grouping and totally arrhythmic "breath-groups." | Prevalence of the same relatively short rhythm, grouped by pauses, clear increments. | Contrast of short upper versus long lower tones, and of groups of equal versus patterned rhythm. | Contrast of long initial versus short last tones, and of free "verbal" versus patterned "musical" rhythm. |
| | 6. Meter | Most irregular, often totally "ametric," a dragging feel. | Can contain regular fragments, usually iambic. | Mostly regular (spondaic). | Usually irregular (loose, free). | Irregular, as if always changing iambic-trochaic. |
| | 7. Articulation | Absolute dominance of legato. Frequent and long caesuras between phrases that usually end on long tenuto tone and a descending glide. | Contrasted groups of staccato on ascending leaps and legato on descending steps. Frequent ending on a tenuto glide. | Prevalent non-legato provides ease and clarity of recognition for each of the imitations. | Contrasted groups of staccato for ascending leaps and legato for steps as well as descending leaps. Sometimes tenuto endings. | Syllables within a word usually are legato, while words or vocables are separated by pauses. |
| 3. Amplitude | 8. Dynamics | The loudest type. Mid-distance kula phrases are shaped diminuendo, while each long-distance kula sentence is shaped wavelike. | Rather intense dynamics with contrasts between the wavelike shape and the opposition of loud high versus soft low tones. | Only minor dynamic changes within mostly loud levels, copying the typical dynamic envelop of a typical animal call. | Moderate contrast of softer high tones and louder lower tones, in overall soft dynamics. | The softest type, yet with sudden accents, falling on a single syllable in those words that are marked by a zigzag leap. |
| 4. Timbre | 9. Register | Single, very high register—the longer the distance, the brighter the tonal quality (piercing or shrilling). Fixed high larynx position, constant brightness. | Multiple registers: shrieking for the highest tones, shrilling (kula-like) for high, "casual" for low tones. Variable larynx position. Recoloring of the tones within a motif. | Single register for each onomatopoeic imitation, usually broadband—in contrast to the narrow-band kula. Usually high larynx. Overall dullness. | Contrast of 2 registers: head-voice (kula-like) for high tones and throat or chest singing voice for low tones. Variable larynx position. Recoloring of the tones within a calling motif. | Contrast of 2 registers: head-voice (kula-like) for one syllable and normal speaking voice for the rest. Variable larynx position. |
| | 10. Harmonicity, attack and vibrato | Clear harmonics (akin to pure tones), ascending portamento attack and descending termination (in longer kulas), minimal vibrato (only to embellish a tone such as a trill). | Clear harmonics (akin to kula) for the highest and longest tones only, ascending and descending portamento, no vibrato. | Prevalence of non-periodic spectrum, harsh, noisy sound, little voicing (only if present in the imitated model), no portamento or vibrato. | Clear and rich harmonics for the lowest and longest tones only (with some vibrato), ascending and descending portamento for leaps only. | Prevalence of non-periodic spectrum, as in speech, with pronounced frequency modulation and noise. No vibrato. Leaps engage portamento. |

*Ten AEs (in rows) are used in five types of phrases (in columns), each characterized by a unique combination of AE patterns, the most distinctive of which are pitch, rhythm, articulation, dynamics, and register. Each type also is distinguished by its semantic specialization: kula—safety signal for grazing, exclamation calls—social "grooming talk," onomatopoeia—playful teasing, parlando—commanding and convincing, and motherese—endearing and trusting. Except for the long-distance kula (whose sentences can reach up to 15 s), all other types are quite brief (usually, 0.3–2 s) and are intermixed with the same-distance or shorter-distance motif types (i.e., kula phrases can be included in motherese recitative, but motherese cannot be included in kula). The maximal distance squeezes the ambitus into an octave confined to a single highest register. This compresses degrees into steps of a smaller or a longer intervallic size, depending on their phrasal position. Climactic step tends to constitute the interval of major 2nd, whereas cadential step—of minor 2nd, to emphasize and facilitate resolution of tension ("major" for a peak in tension, "minor" for relaxation). The framework of a breath cycle sets the basis for traditional association of major with happiness (climax = inspiration = maximal power) and of minor with sadness (cadence = expiration = collapse). Mid-distance kula transposes heptatonic structures to lower registers, fitting them into a tetrachord (pentachord, if a climactic motif is added). Octave equivalence secures heptatony. Closer distances enable alterations, flattening or sharpening of the unstable degrees, and timbral recoloring. Stacking phrases of contrasting TO and semantic values, learnable by humans and domesticated animals, generates the SFTO. See a fuller version of this table in Appendix 2 "A Comparative Structural Analysis of Musograms," in* ***Supplementary Material****.*

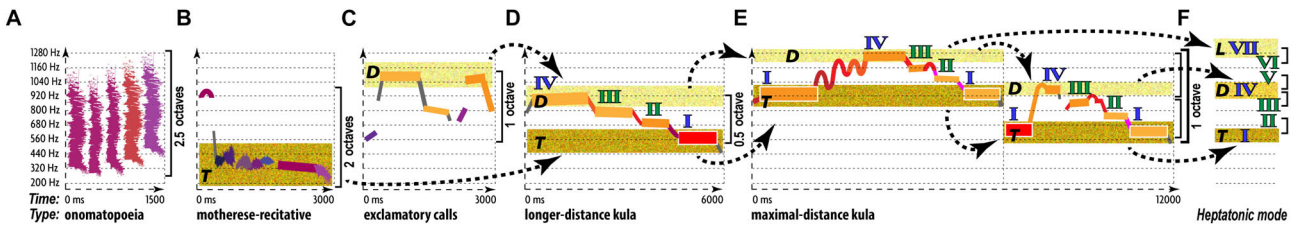



**FIGURE 8 |** Genesis of SFTO in the vocal tradition of kulning. The set of six panels shows five excerpts representing five different vocalization types from **Figure 7**. They are placed on the same frequency grid to demonstrate how the registral position of phrasal tones evolves into a frequency range used to define a degree in a musical mode. Thin dashed vertical line indicates the phrasal ends. Thick curved dashed arrows show the genesis of "tonic" (principal stable) and "dominant" (principal unstable) degrees, eventually shaping a heptatonic mode. **(A) Mid-distance onomatopoeia (barking)**. This is the closest to ACs. A phrase repeats the same wideband aperiodic signal whose most intense part of the spectrum spreads over ≈2.5 octaves. **(B) Close distance motherese/recitative**. Low-register tones in such phrases tend to fall within the same narrow range of 250–290 Hz (257 c), marked by the darker grainy filling. Frequently repeated voiced vowels effectively refine the tuning of the "recitative tone" that adopts a tonic function ("T") established by the common terminations of phrases. **(C) Mid-distance exclamatory calls**. High-register "shrieking" tones in such phrases are squeezed in a twice narrower range, marked by a lighter filling. These shrieking tones complement low tones in **(B)** in providing reference for pitch changes. Such tones prototype the "dominant" melodic function ("D"). Tones that fall in this register become imperfect "anchors" subdued to "tonic"—requiring a descending melodic "resolution" after them. **(D) Longer-distance kula**. Registral ranges of both "tonic" and "dominant" are further compressed into "degrees" of a simple 4-degree musical mode. Colored Roman numerals use blue for anchored tones, and green—for supporting tones (passing or auxiliary). Tonic function (stability) is shaped by the lowest degree terminating a phrase, whereas dominant function (instability)—by the highest degree initiating a phrase. This transformation is fueled by frequent stitching of **(B)**, **(C)**, and **(D)** phrases within the same musicking session as the distance changes. **(E) Longest-distance kula**. This type doubles the TO structure of the shorter distance kula—indicated by two thin vertical brackets encapsulated by one thick bracket. A similar tetrachordal structure is reproduced above the base-tetrachord. Both tetrachords are conjoined: the lowest stable degree ("T") of an upper tetrachord becomes the highest unstable degree ("D") of the lower tetrachord as kula descends from its opening phrase toward lower phrases, terminated by the lowest permanent "tonic." Repeated use of such complex structure (common at distances over 1 km) is likely to turn it into a modal framework for the entire kulning, encompassing all its phrasal types. **(F) Heptatonic mode in complex kulas**. Frequent modulations between the conjoined tetrachords integrate both tetrachords into a single complex 2-tetrachordal mode with three axial degrees: the lowest *I*—a permanent tonic ("T"), the middle *IV*—an alternative temporary anchor that requires resolution ("D"), and the highest *VII*—a permanent unstable anchor, used to initiate sentences and/or build a climax—i.e., the "leading tone" ("L") that always leads to more stable anchors (perfect and/or imperfect). These axial degrees enclose supplementary degrees, each of which is bound to the closest anchor, forming pairs.

herding music promotes monodic application: herding demands hours of solitary interaction with animals, ideal for testing their response to music-making.

## CONCLUSION

Homo heidelbergensis was already anatomically capable of practicing proto-music which was most probably isophonic, lacking the combined coding of pitch/rhythm—without which conventionalization of the semiotically functional melody-making was hardly possible. Isophony supports only group communication of zero- and first-order intentionality, limited and conditioned by the genetically embedded instinctive responses to isophonic formula. Learning of multi-factorial particulate expression and second-order intentionality requires monophonic production. AE's pattern becomes fully semiotic only when many senders/receivers remember it as the bearer of the same semantic value that connotes a certain affective state—"binding hearer to speaker" through "tying of some social sentiment" (Wallin, 1991, 420).

Emotional contagion is possible in isophonic signals, but it is primed to a single most salient AE—provided all communicators share the necessary neuro-anatomical substrates (Snowdon et al., 2015). Harmony, meter, texture, and form are not supported by non-human brains; neither is a premediated "construction" of an intended message. Animal interpretation of auditory signals is inherently circumstantial—determined by the signaling context (Zuberbühler, 2017). Therefore, human music is often "misunderstood" by animals, requiring music's "translation" into animal's "sonic templates of recognition" (Snowdon and Teie, 2013).

For ACs to evolve into music, a repertory of patterns of AEs had to be extracted from proto-musicking practice and abstracted into elemental signs to continuously inform someone(s) of the communicator's affective state, intentions, and needs. Such use emerged in communication with domesticated dogs, thereafter, adapted for herding. Hunting/gathering does not demand such communication. Instead, it prioritizes *collective* collaboration: bringing participants emotionally "in-tune," binding them into a group to increase one's powers. Such use makes sense in situations of using loud complex sounds while hunting large prey and repelling human predators in open savannah space (Jordania, 2011). Large groups of big-game foragers tend to prioritize *collective* music-making over *personal*, confining the latter to pre-pubertal age, like Aka Pygmies (Rouget, 2011). Homo probably exported isophonic proto-music from Africa to Europe.

The last Glacial Maximum greatly reduced the European population by the Gravettian—until the Magdalenian repopulation (Maier, 2017) enabled the rise of symbolic cultures (Kozłowski, 2015) and ethnolinguistic genesis (Zilhão, 2014). Low-density foraging groups usually form alliances, cemented by linguistic commonalities and intermarriage (Marlowe, 2005). Music surpasses language in its bonding capacities (Nakata and Trehub, 2004). Gravettian proto-music must have adjusted isophony for new cultural applications,

especially religious. Smaller groups generate a smaller sonic "jumble," facilitating the recognition of specific musical elements. Smaller groups also promote honesty in communication (Richerson and Boyd, 2005). Honest musical expression enables and validates the person-to-person musical communication. This opens doors to the cultural development of a motherese communicative model. Small groups are likely to promote motherese-like duetic and babbling-like solitary music-making. Thus, collective music-making is exceedingly rare in Northern Siberia (Alekseyev, 1967) which has always remained underpopulated (Sikora et al., 2019)—closely resembling life in glaciated Europe.

Motherese talk, lullabies, onomatopoeia, and instinctive utterances supplied the initial material for the formation of bi-specific SFTO. Changes in distance while continuously communicating with the herd put in place the musical modes. The closest distance promotes low-register monotony, middle distance—high-register monotony, long distance—tetrachord-based tonicity, and maximal distance—conjunct pentachord/tetrachord octave-equivalent modes with dominant-tonic functionality. Monotony increases the tuning accuracy of anchor-tones, firstly defining principal degrees (tonic, supertonic, dominant), and then additional unstable degrees (Alekseyev, 1976). Characteristic modal intonations of different phrasal styles and varying position within a breathing cycle charge modal degrees with specific functionality, which directs the formation of semantic values for each of the common modal intonations. This triggers the process of modal evolution as outlined by Beliaev (1963) and elaborated by Nikolsky (2015a, 2016).

Nordic kulning is probably a vestige of an archaic cattle-oriented "domestication language" which descended from yodel—accompanying the northerly spread of Indo-European languages throughout Europe. Other Eurasian domestication languages accompanied the spread of the Uralic and Turkic language families, and were optimized, respectively, for reindeer and horse. Rémy Dor cross-analyzed vocalizations/whistles of herders speaking 20 Turkic languages, from Anatolia to Yakutia, and inferred their syntactic organization (Dor, 2005), identifying their common utterances (Dor, 1993). Like Wallin and Alekseyev, Dor too found continuity between vocalizations of hunters and herders: "somatotropic" vocalizations, designed to make the prey come closer, evolved into "fetch" or "home-return" calls, while "somatofugal" vocalizations evolved into "stop" calls to repel predators. The new class of "somatoneutral" vocalizations emerged in order to keep an animal at a constant distance (like safety-call kula). Strong biological foundation of this distance-governed communication made it well-conserved—practically indestructible—unlike languages or music systems (Dor, 2008).

Domestication languages could underlie modern languages and musics, as traditional beliefs suggest. Swedish rural informants considered kulning an ancient "language" (Moberg, 1971, 145). And on the opposite end of Eurasia, Mongolian herders believe that their music-making is derivative of the "large language," superior to human language and designed to communicate with animals, nature, and spirits (Pegg, 2001, 235). Altaic *xöömii* most likely constitutes yet another "domestication language."

Capacity to simultaneously control numerous AEs and second-order intentionality enabled humans to create a heterospecific semiotic system of communicating desirable affective states, which gave humans control over domestic animals, resolved human sustenance needs, and put in place music as we know it. The semiotically functional tonal organization that distinguishes music from speech might have emerged no earlier than during the Neolithic "revolution" as a result of forging new conventions of human-to-animal vocal communication.

## DIRECTIONS FOR FUTURE RESEARCH

Comparative examination of human-to-animal signaling for different domesticate animals across different geographic regions can confirm whether the paradigm of "musical domestication language," divisible in "dialects" and integrable into "language families," is applicable here.

Collecting a database of patterns of human-to-animal communication would be analogous to building a lexicon of a newly discovered natural human language or to establishing a stock of typical idioms in the musical communication within a novel musical culture. Once established, such database can be statistically analyzed and cross-examined in relation to other databases, e.g., of emotional expressions in music. This could substantiate or invalidate my conclusions.

The perception of specific elements and patterns of human-to-animal communication by humans and animals can be experimentally tested. This could identify syntactic and pragmatic rules that cannot be assessed by acoustic analysis alone. Together, both approaches can evaluate semiotic efficacy of TO in pastoral signaling. This, in turn, can establish whether introduction of herding communication during the Neolithic Revolution was capable of generating SFTO in music to make it break away from the basics of animal communication.

Experimental archaeo-ethnomusicology could provide yet another way of verifying this hypothesis. Members of isolated tribes that maintain a hunter/gatherer lifestyle and use no domestic animals can be introduced to domestic animals and "taught" to use music-like signals to command them. Their progress can be analyzed and compared to patterns of conspecific acquisition of music skills by human infants, as well as to the available archaeological, genetic, and paleo-physiological data.

## AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

## ACKNOWLEDGMENTS

I am grateful to CT and MR for reviewing the manuscript for this paper, and to Sheila Bazleh for copy-editing it. My special thanks to Leonid Perlovsky, Steven Brown, Piotr Podlipniak, Leon Crickmore, Theodor Levin, Margarita Mazo, and Philipp

Tagg for their critical input in relation to matters of semiotics of music, and to Isaly Zemtsovsky, Eduard Alekseyev, and Frank Scherbaum for reviewing my approach to modal analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01358/full#supplementary-material

**DATA SHEET S1 |** Appendix 1 – A new method of modal multifactorial analysis of tonal organization in music and music-like sounds.
This technical paper contains instructions for identifying the tonal organization in a music work, a music-like vocalization (e.g., infant's babbling) or music-like animal signals (e.g., bird's song) – including sounds that are indefinite or modulating in pitch.

**DATA SHEET S2 |** Appendix 2 – A comparative structural analysis of musograms used in **Figures 3**, **4**, **7** of this article.
This document contains a comprehensive analysis of the characteristic traits of tonal organization in the examples of human musical communication, animal vocal communication, and bi-specific communication between domestic animals and their human guardians.

## REFERENCES


Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *J. Soc. Biol. Struct.* 12, 1–13. doi: 10.1016/0140-1750(89)90015-8

Addessi, A. R., and Caterina, R. (2000). Perceptual musical analysis: segmentation and perception of tension. *Music. Sci.* 4, 31–54. doi: 10.1177/102986490000400102

Adorno, T. W. (1942). *On Popular Music*. Frankfurt am Main: Institute of Social Research.

Agawu, V. K. (2003). *Representing African Music: Postcolonial Notes, Queries, Positions*. London: Routledge.

Agawu, K. (2004). How we got out of analysis, and how to get back in again. *Music Analysis* 23, 267–286. doi: 10.1111/j.0262-5245.2004.00204.x

Ahlbäck, S. (2007). *The Tonality of Older Swedish Folk Music*. Stockholm: Udda Toner.

Aiello, L. C. (1998). "The foundations of human language," in *The Origin and Diversification of Language*, eds N. G. Jablonski and L. C. Aiello (San Francisco, CA: California Academy of Sciences), 21–34.

Aikio, A. (2006). New and old samoyed etymologies. part-2. *Finnisch Ugrische Forschungen* 59, 9–34.

Aksyonov, A. N. (1964). *Tuvan Folk Music: Materials and Studies [Тувинская народная музыка: материалы и исследования]*. Moscow: Muzyka.

Alekseyev, E. Y. (1967). Is there polyphony among the Yakuts? [Есть ли у якутов многоголосие]. *Sovetskaya Muzyka* 5, 97–105.

Alekseyev, E. Y. (1976). *Problems in the Genesis of Musical Mode (on the example of Yakut folksong): analysis [Проблемы формирования лада (на материале якутской народной песни): исследование]*. Moscow: Muzyka.

Alekseyev, E. Y. (1986). *Musical Intonation in the Earliest Forms of Folklore. The Aspect of Pitch [Раннефольклорное интонирование: звуковысотный аспект]*. Moscow: Soviet Composer.

Alekseyev, E. Y. (1995). "The pragmatics, hystrionics and rituals in sound imitations of peoples of Northern Siberia [Прагматическое, игровое, ритуальное в звукоподражаниях народов Севера Сибири]," in *Voice and Ritual. [Голос и ритуал]*, eds E. Dorokhova, N. Zhulanova, and O. Pashina (Moscow: The State Institute of Art Studies at the Ministry of Culture of the Russian Federation), 33–35.

Alekseyev, E. Y., and Nikolayeva, N. (1981). *Models of Yakut vocal folklore [Образцы якутского песенного фольклора]*. Yakutsk: Academy of Sciences of the USSR.

Altenmüller, E., Kopiez, R., and Grewe, O. (2013a). "A contribution to the evolutionary basis of music: lessons from the chill response," in *Evolution of Emotional Communication*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford: Oxford University Press), 313–336. doi: 10.1093/acprof:oso/9780199583560.003.0019

Altenmüller, E., Kopiez, R., and Grewe, O. (2013b). "Strong emotions in music: are they an evolutionary adaptation?," in *Sound - Perception - Performance*, ed. R. Bader (Heidelberg: Springer), 131–156. doi: 10.1007/978-3-319-00107-4_5

Altenmüller, E., Schmidt, S., and Zimmermann, E. (eds). (2013). "A cross-taxa concept of emotion in acoustic communication: an ethological perspective," in *Evolution of Emotional Communication*, (Oxford, UK: Oxford University Press), 339–356. doi: 10.1093/acprof:oso/9780199583560.003.0020

Alvarez-Pereyre, F., and Arom, S. (1993). Ethnomusicology and the emic/etic issue. *World Music* 35, 7–33.

Amaya, K., Bruderlin, A., and Calvert, T. (1996). "Emotion from Motion," in *Proceedings of Graphics Interface '96 Toronto, Ontario, Canada: 22 - 24 May 1996*, eds R. Bartels and W. A. Davis (Toronto: Canadian Human-Computer Communications Society), 222–229.

Ambrose, S. H. (2001). Paleolithic technology and human evolution. *Science* 291, 1748–1753. doi: 10.1126/science.1059487

Ambrose, S. H. (2010). Coevolution of composite-tool technology, constructive memory, and language. *Curr. Anthropol.* 51, S135–S147. doi: 10.1086/650296

Anthony, D. W. (1995). Horse, wagon & chariot: indo-European languages and archaeology. *Antiquity* 69, 554–565. doi: 10.1017/S0003598X00081941

Anthony, D. W., and Ringe, D. (2015). The indo-european homeland from linguistic and archaeological perspectives. *Annu. Rev. Linguist.* 1, 199–219. doi: 10.1146/annurev-linguist-030514-124812

Aranovsky, M. G. (1991). *Syntactic Structure of a Melody. Research [Синтаксическая структура мелодий. Исследование]*. Moscow: Muzyka.

Aranovsky, M. G. (1998). *Musical Text. Structure and Properties [Музыкальный текст. Структура и свойства]*. Moscow: Kompozitor.

Armstrong Oma, K. (2010). Between trust and domination: social contracts between humans and animals. *World Archaeol.* 42, 175–187. doi: 10.1080/00438241003672724

Armstrong Oma, K. (2013). Human-animal meeting points: use of space in the household arena in past societies. *Soc. Anim.* 21, 162–177. doi: 10.1163/15685306-12341300

Arom, S. (2004). *African Polyphony and Polyrhythm: Musical Structure and Methodology*. Cambridge, MA: Cambridge University Press.

Arom, S. (2010). Corroborating external observation by cognitive data in the description and modelling of traditional music. *Music. Sci.* 14, 295–306. doi: 10.1177/10298649100140S216

Arom, S., and Voisin, F. (1997). "Theory and technology in African music," in *The Garland encyclopedia of world music: Africa*, ed. R. M. Stone (New York, NY: Garland Publishing), 254–270.

Aronsson, K. -Å (1991). *Forest Reindeer Herding A.D. 1-1800: An Archaeological and Palaeoecological Study in Northern Sweden*. Umeå: Dept. of Archaeology, University of Umeå.

Atkinson, Q. D., and Gray, R. D. (2006). "How old is the Indo-European language family? Illumination or more moths to the flame?," in *Phylogenetic Methods and the Prehistory of Languages*, eds P. Forster and C. Renfrew (Cambridge: McDonald Institute for Archaeological Research), 91–109.

Atkinson, Q. D., Nicholls, G., Welch, D., and Gray, R. D. (2005). From words to dates: water into wine, mathemagic or phylogenetic inference? *Trans. Philol. Soc.* 103, 193–219. doi: 10.1111/j.1467-968X.2005.00151.x

August, P. V., and Anderson, J. G. T. (1987). Mammal sounds and motivation-structural rules: a test of the hypothesis. *J. Mammal.* 68, 1–9. doi: 10.2307/1381039

Axelsson, E., Ratnakumar, A., Arendt, M. L., Maqbool, K., Webster, M. T., Perloski, M., et al. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360–364. doi: 10.1038/nature11837

Bader, R. (ed.) (2018). *Springer Handbook of Systematic Musicology*. Berlin: Springer-Verlag.

Balari, S., Benítez-Burraco, A., Camps, M., Longa, V. M., Lorenzo, G., and Uriagereka, J. (2011). The archaeological record speaks: bridging anthropology and linguistics. *Int. J. Evol. Biol.* 2011:382679. doi: 10.4061/2011/382679

Banshchikov, G. (1997). *The Rules of Functional Instrumentation [Законы функциональной инструментовки]*. Sankt-Petersburg: Kompozitor.

Bartel, D. (1997). *Musica Poetica: Musical-Rhetorical Figures in German Baroque Music*. Lincoln: University of Nebraska Press.

Baumann, M. (1993). Listening as an emic/ethic process in the context of observation and inquiry. *World Music* 35, 34–62.

Baumann, M. P. (1976). *Musikfolklore und Musikfolklorismus*. Winterthur: Amadeus Press.

Becker, J. (1986). Is western art music superior? *Music. Q.* 72, 341–359. doi: 10.1093/mq/LXXII.3.341

Beliaev, V. (1963). The Formation of folk modal systems. *J. Int. Folk Music Council* 15, 4–9. doi: 10.2307/836227

Beliayev, V. M. (1990a). "Modal systems in the traditional music of the USSR [Ладовые системы в музыке народов СССР]," in *Viktor Mikhailovich Beliayev [Виктор Михай лович Беляев]*, ed. I. Travkina (Moscow: Soviet Composer), 223–377.

Beliayev, V. M. (1990b). *Viktor Mikhailovich Beliayev [Виктор Михайлович Беляев]*. Moscow: Soviet Composer.

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., and Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proc. R. Soc. B Biol. Sci.* 275, 473–481. doi: 10.1098/rspb.2007.1460

Bellwood, P. S. (2008). "Archaeology and the origins of language families," in *Handbook of Archaeological Theories*, eds R. A. Bentley, H. D. G. Maschner, and C. Chippindale (Lanham, MD: Altamira Press), 225–244.

Benjamin, T., Horvit, M. M., and Nelson, R. (2015). *Techniques and Materials of MUSIC: From the Common Practice Period through the Twentieth Century*. Stanford, CA: Cengage Learning.

Benson, D. J. (2007). *Music: A Mathematical Offering*. Cambridge: Cambridge University Press.

Benward, B., and Saker, M. N. (2009). *Music in Theory and Practice*. New York, NY: McGraw-Hill.

Berliner, P. F. (1993). *The Soul of the Mbira: Music and Traditions of the Shona People of Zimbabwe*. Chicago, IL: University of Chicago Press.

Berndt, A., and Hähnel, T. (2010). "Modelling musical dynamics," in *Proceedings of the 5th Audio Mostly Conference: A Conference on Interaction with Sound, Piteå, Sweden — September 15 - 17, 2010*, ed. K. Brandenburg (Ilmenau: Fraunhofer Institute for Digital Media Technology), 1–8.

Berry, W. T. (1987). *Structural Functions in Music*. New York, NY: Dover Publications.

Bersaglieri, T., Sabeti, P. C., Patterson, N., Vanderploeg, T., Schaffner, S. F., Drake, J. A., et al. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* 74, 1111–1120. doi: 10.1086/421051 doi: 10.1086/421051

Berthon, R. (2014). Past, current and future contribution of zooarchaeology to the knowledge of the neolithic and chalcolithic cultures in south caucasus. *Stud. Caucasian Archaeol.* 2, 4–30.

Blacking, J. (1967). *Venda Children's Songs: A Study in Ethnomusicological Analysis*. Chicago: University of Chicago Press.

Blacking, J. (1974). *How Musical is Man?*. Seattle, WA: University of Washington Press.

Blake, E. C. (2011). *Stone "Tools" as Portable Sound-Producing Objects in Upper Palaeolithic Contexts: The Application of an Experimental Study*. Dissertation, University of Cambridge, The Department of Archaeology, Cambridge.

Bläuer, A., and Kantanen, J. (2013). Transition from hunting to animal husbandry in Southern, Western and Eastern Finland: new dated osteological evidence. *J. Archaeol. Sci.* 40, 1646–1666. doi: 10.1016/j.jas.2012.10.033 doi: 10.1016/j.jas.2012.10.033

Blench, R. (2013). Methods and results in the reconstruction of music history in Africa and a case study of instrumental polyphony. *Azania Archaeol. Res. Afr.* 48, 31–64. doi: 10.1080/0067270X.2013.771016

Bobrovsky, V. P. (1978). *The Functional Basics of Musical Form [Функциональные основы музыкальной формы ]*. Moscow: Muzyka.

Bogucki, P. I. (1988). *Forest Farmers and Stockherders: Early Agriculture and Its Consequences in North-Central Europe*. Cambridge, MA: Cambridge University Press.

Boivin, N. (2004). Landscape and cosmology in the South Indian Neolithic: new perspectives on the Deccan Ashmounds. *Camb. Archaeol. J.* 14, 235–257. doi: 10.1017/S0959774304000150

Boivin, N., Brumm, A., Lewis, H., Robinson, D., and Korisettar, R. (2007). Sensual, material, and technological understanding: exploring prehistoric soundscapes in south India. *J. R. Anthropol. Institute* 13, 267–294. doi: 10.1111/j.1467-9655.2007.00428.x

Bonfeld, M. S. (2006). *Music - Language, Speech, Thinking: Experience of Systematic Research of Musical Art [Музыка - язык, речь, мышление: опыт системного исследования музыкального искусства]*. Sankt-Petersburg: Kompozitor.

Bordzhanova, T. G. (2007). *Ritual Poetry of the Kalmyks. The system of genres, poetics. [Обрядовая поэзия калмыков. Система жанров, поэтика]*. Elista: Kalmyk Book Publishers.

Braudo, I. A. (1961). *Articulation. On pronounciation of melody [Артикуляция. О произношений мелодий.]*. Leningrad: State musical publishing.

Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge MA: MIT Press.

Bresin, R., and Friberg, A. (2011). Emotion rendering in music: range and characteristic values of seven musical variables. *Cortex* 47, 1068–1081. doi: 10.1016/j.cortex.2011.05.009

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *J. Zool.* 288, 1–20. doi: 10.1111/j.1469-7998.2012.00920.x

Brodsky, I. A. (1976). "On research of music of the ethnicities of North of the Russian Federation [ К изучению музыки народов Севера РСФСР]," in *Traditional and Contemporary Folk Music Art. Collection of Essays. [Традиционное и современное народное музыкальное искусство. Сборник мрудое]*, ed. B. Yefinmenkova (Moscow: State Music Pedagogical Institute named after Gnesin), 244–257.

Brooks, R. R. R., and Wakankar, V. S. (1976). *Stone Age Painting in India*. London: Yale University Press.

Brown, K. S., Marean, C. W., Herries, A. I. R., Jacobs, Z., Tribolo, C., Braun, D., et al. (2009). Fire as an engineering tool of early modern humans. *Science* 325, 859–862. doi: 10.1126/science.1175028

Brown, N. W. (1964). The sanctity of the cow in hinduism. *Econ. Wkly.* 1, 245–256.

Brown, S. (2000). "The 'Musilanguage' model of language evolution," in *The Origins of Music*, eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 271–300.

Brown, S. (2007). Contagious heterophony: a new theory about the origins of music. *Music. Sci.* 11, 3–26. doi: 10.1177/10298649070110 0101 doi: 10.1177/102986490701100101

Brown, S. (2017). A joint prosodic origin of language and music. *Front. Psychol.* 8:1894. doi: 10.3389/fpsyg.2017.01894

Brudzynski, S. M. (2013). "Vocalizations as indicators of emotional states in rats and cats," in *Evolution of Emotional Communication*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford, UK: Oxford University Press), 75–91. doi: 10.1093/acprof:oso/9780199583560.003.0005

Burger, J., Kirchner, M., Bramanti, B., Haak, W., and Thomas, M. G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. *Proc. Natl. Acad. Sci. U.S.A.* 104, 3736–3741. doi: 10.1073/pnas.0607187104

Bytchkov, Y. N. (1987). On Systemic Nature of Modal Organization in Music [ О системном характере ладовой организации в музыке]. Dissertation, Russian Academy of Music named after Gnessin, The Department of Music theory, Moscow. doi: 10.1073/pnas.0607187104

Bytchkov, Y. N. (1997). *On Dialectics of Formation and Unfolding of Mode [О диалектике становления и развертывания лада]*. Moscow: Russian Academy of Music named after Gnessin.

Cabouret, M. (1984). *La vie Pastorale Dans les Montagnes et les Forêts de la Péninsule Scandinave*. Dissertation, Université de Paris-Sorbonne, The Department of Geography, Lille.

Cambouropoulos, E. (2008). Voice and stream: perceptual and computational modeling of voice separation. *Music Percept.* 26, 75–94. doi: 10.1525/mp.2008.26.1.75

Cambouropoulos, E. (2010). The musical surface: challenging basic assumptions. *Music. Sci.* 14, 131–147. doi: 10.1177/10298649100140S209

Campbell, A. K., Waud, J. P., and Matthews, S. B. (2005). The molecular basis of lactose intolerance. *Sci. Prog.* 88, 157–202. doi: 10.3184/003685005783238408

Campbell, J. (2017). *The Masks of God. Occidental Mythology*. San Anselmo, CA: Joseph Campbell Foundation.

Camps, M., and Uriagereka, J. (2006). "The Gordian knot of linguistic fossils," in *The Biolinguistic Turn: Issues on Language and Biology*, eds J. Rosselloì and J. Martiìn (Barcelona: University of Barcelona), 34–65.

Caramelli, D. (2006). The origins of domesticated cattle. *Hum. Evol.* 21, 107–122. doi: 10.1007/s11598-006-9013

Chang, H.-W., and Trehub, S. E. (1977). Auditory processing of relational information by young infants. *J. Exp. Child Psychol.* 24, 324–331. doi: 10.1016/0022-0965(77)90010-8

Chang, W., Cathcart, C., Hall, D., and Garrett, A. (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91, 194–244. doi: 10.1353/lan.2015.0005

Charlton, D. (2009). *Opera in The age of Rousseau: Music, Confrontation, Realism*. Cambridge, MA: Cambridge University Press.

Chase, P. (2006). *The Emergence of Culture: The Evolution of a Uniquely Human Way of Life*. New York, NY: Springer.

Cheape, H. (1996). Shielings in the highlands and islands of scotland: prehistory to the present. *Folk Life* 35, 7–24. doi: 10.1179/043087796798254498

Cheng, J. (2009). A review of early dynastic iii music: man's animal call. *J. Near East. Stud.* 68, 163–178. doi: 10.1086/613988

Chew, G. (2001). *Articulation And Phrasing*. Oxford: Oxford University Press.

Christensen, T. (ed.) (2008). *The Cambridge History of Western Music Theory*. Cambridge, MA: Cambridge University Press.

Clarke, E. F. (2007). "Rhythm and timing in music," in *The Psychology of Music*, (Cambridge, MA: Academic Press), 473–500. doi: 10.1016/b978-012213564-4/50014-7

Clayton, M. R. L. (2000). *Time in Indian Music: Rhythm, Metre, and Form in North Indian Rag Performance*. Oxford, UK: Oxford University Press.

Conard, N. J., Malina, M., and Münzel, S. (2009). New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460, 737–740. doi: 10.1038/nature08169

Cook, N. D. (2002). *Tone of Voice and Mind: The Connections between Intonation, Emotion, Cognition and Consciousness*. Amsterdam: John Benjamins Publishing.

Cooke, D. (1959). *The Language of Music*. London: Oxford University Press.

Cramp, L. J. E., Evershed, R. P., Lavento, M., Halinen, P., Mannermaa, K., Oinonen, M., et al. (2014). Neolithic dairy farming at the extreme of agriculture in northern Europe. *Proc. R. Soc. B Biol. Sci.* 281:20140819. doi: 10.1098/rspb.2014.0819

Crickmore, L. (2009). "The tonal systems of mesopotamia and ancient greece: some similarities and differences," in *The Archaeomusicological Review of the Ancient Near East*, eds R. Dumbrill and M. Marcetteau (London: Academic Press), 1–16.

Cross, I. (1999). "Is music the most important thing we ever did? music, development and evolution," in *Music, Mind and Science*, ed. S. W. Yi (Seoul: Seoul National University Press), 10–39.

Cross, I., and Watson, A. (2006). "Acoustics and the human experience of socially-organized sound," in *Archaeoacoustics*, eds C. Scarre and G. Lawson (Cambridge, MA: McDonald Institute for Archaeological Research), 107–116.

Crossley-Holland, P. (1967). "Form and style in Tibetan folksong melody," in *Jahrbuch für musikalische Volks- und Völkerkunde Band 3*, ed. F. Bose (Berlin: Walter de Gruyter), 9–69. doi: 10.1515/9783111448503-002

Curry, A. (2013). The milk revolution. *Nature* 500, 20–22. doi: 10.1038/500020a

Cynx, J. (1990). Experimental determination of a unit of song production in the zebra finch (*Taeniopygia guttata*). *J. Comp. Psychol.* 104, 3–10. doi: 10.1037/0735-7036.104.1.3

Dahlhaus, C. (1989). *Nineteenth-Century Music*. Berkeley, CA: University of California Press.

Dasen, P. R. (2012). "Emics and Etic in Cross-Cultural Psychology Towards a Convergence in the Study of Cognitive Styles," in *Proceedings of the 4th Africa Region Conference of the IACCP, University of Buea, Cameroun, Aug. 1-8, 2009*, eds T. M. S. Tchombe and M. Fülöp (Buea: University of Buea), 55–73.

Davidson, M. (1970). Some music for the Lala Kankobele. *Afr. Music J. Afr. Music Soc.* 4, 103–113. doi: 10.21504/amj.v4i4.1685

de Götzen, A. (2004). "The sounding gesture: an overview," in *Proceedings of the 7th International Conference on Digital Audio Effects, DAFx 04: Naples, October 5 - 8, 2004*, ed. G. Evangelista (Naples: Università degli studi di Napoli, Dipartimento di scienze fisiche), 5–10.

Dean, R. T., Bailes, F., and Schubert, E. (2011). Acoustic intensity causes perceived changes in arousal levels in music: an experimental investigation. *PLoS One* 6:e18591. doi: 10.1371/journal.pone.0018591

Demany, L., Mckenzie, B., and Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature* 266, 718–719. doi: 10.1038/266718a0

Dennett, D. C. (1983). Intentional systems in cognitive ethology: the "Panglossian paradigm" defended. *Behav. Brain Sci.* 6, 343–355. doi: 10.1017/S0140525X00016393

d'Errico, F., and Stringer, C. B. (2011). Evolution, revolution or saltation scenario for the emergence of modern cultures. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1060–1069. doi: 10.1098/rstb.2010.0340

D'Errico, F., Villa, P., Llona, A. C. P., and Idarraga, R. R. (1998). A middle palaeolithic origin of music? Using cave-bear bone accumlations to assess the Divje. *Antiquity* 72:65. doi: 10.1017/s0003598x00086282

Deutsch, D. (2013). "The processing of pitch combinations," in *Psychology of Music*, ed. D. Deutsch (New York: Academic Press), 249–325. doi: 10.1016/b978-0-12-381460-9.00007-9

Diamond, J., and Bellwood, P. S. (2003). Farmers and their languages: the first expansions. *Science* 300, 597–603. doi: 10.1126/science.1078208

Diedrich, C. G. (2015). 'Neanderthal bone flutes': simply products of Ice Age spotted hyena scavenging activities on cave bear cubs in European cave bear dens. *R. Soc. Open Sci.* 2, 140022–140022. doi: 10.1098/rsos.140022

Dissanayake, E. (2005). "Ritual and ritualization: musical means of conveying and shaping emotion in humans and other animals," in *Music and Manipulation: On the Social Uses and Social Control of Music*, eds S. Brown and U. Volgsten (New York, NY: Berghahn Books), 31–56.

Dissanayake, E. (2008). If music is the food of love, what about survival and reproductive success? *Music. Sci.* 12, 169–195. doi: 10.1177/1029864908012001081

Dobzhanskaya, O. (2011). "Ritual musical folklore of Samoyedic peoples as an object of study: main results and research perspectives [обрядовыймузыкальныйфольклорсамодийскихнародовкакобъектизученияосновныерезультаты и перспективыисследования]," in *From Congress to Congress: Proceedings of the Second All-Russian Congress of Folklore Study [Материалывсероссийскогоконгрессафольклористов]*, ed. A. S. Kargin (Moscow: State Republican Center of Russian Folklore), 300–311.

Dobzhanskaya, O. E. (2012). "On the concept of 'music' in relation to archaic and early fokloric cultures [О понятии «музыка» применительно к архаичным и раннефольклорным культурам]," in *XVI Tzarskoselsk Readings [XVI Царскосельские чтения]*, ed. V. Skvortsov (Sankt-Petersburg: Pushkin Leningrad State University), 256–259.

Dobzhanskaya, O. E. (2016). The living has sound; the dead is silent. *Anthropol. Archeol. Eurasia* 55, 7–21. doi: 10.1080/10611959.2016.1263485

Dor, R. (1993). Les Huchements du berger turc: II: du huchement-aux-morts à l'appel des chevaux. *Études Turques Ottomanes* 3, 27–41.

Dor, R. (2005). Les huchements du berger turc III: interpellatifs adressés au gros bétail. *Turcica* 27, 199–222. doi: 10.2143/turc.27.0.2004362

Dor, R. (2008). À l'aube du cri: de l'homme à l'animal avant le partage du monde. *Diogène* 200, 129–139. doi: 10.3917/dio.200.0129

Dorina, M. V. (2004). *Wind and Percussive Instruments of Sayano-Altaian Turks: An Attempt of Historic-Ethnographic Investigation [Духовые и ударные музыкальные инструменты тюрков Саяно-Алтая (Опыт историко-этнографического исследования)]*. Dissertation, Tomsk University, The Department of Ethnography and History of Science, Tomsk.

Dowling, W. J. (1984). Cognitive processes in the perception of art. *Adv. Psychol.* 19, 145–163. doi: 10.1016/S0166-4115(08)62350-X

Drabkin, W. (2001). "Register," in *The New Grove Dictionary of Music and Musicians*, eds S. Sadie, and J. Tyrrell (Oxford: Oxford University Press).

Driscoll, C. A., Macdonald, D. W., and O'Brien, S. J. (2009). From wild animals to domestic pets, an evolutionary view of domestication. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9971–9978. doi: 10.1073/pnas.0901586106

Druzhkova, A. S., Thalmann, O., Trifonov, V. A., Leonard, J. A., Vorobieva, N. V., Ovodov, N. D., et al. (2013). Ancient DNA analysis affirms the canid from altai as a primitive dog. *PLoS One* 8:e57754. doi: 10.1371/journal.pone.0057754

Dumbrill, R. (1998). *The Musicology and Organology of the Ancient Near East*. London: Tadema Press.

Duncan-Kemp, A. M. (1952). *Where Strange Paths Go Down*. Brisbane: W. R. Smith & Paterson.

Dzenzelevskii, I. A. (1984). "Taboos in practice of Carpathian sheep-breaders [Запреты в практике карпат ских овцеводов]," in *Slavic and Balkan folklore: Ethnogenetic commonality and typological parallels [Славянский и балканский фольклор: этногенетическая общность и типологические параллели]*, ed. N. I. Tolstoy (Moscow: Muzyka), 256–277.

Edwards, C. J., Ginja, C., Kantanen, J., Pérez-Pardal, L., Tresset, A., Stock, F., et al. (2011). Dual origins of dairy cattle farming - Evidence from a comprehensive survey of european Y-chromosomal variation. *PLoS One* 6:e15922. doi: 10.1371/journal.pone.0015922

Eerola, T., Friberg, A., and Bresin, R. (2013). Emotional expression in music: contribution, linearity, and additivity of primary musical cues. *Front. Psychol.* 4:487. doi: 10.3389/fpsyg.2013.00487

Eerola, T., Järvinen, T., Louhivuori, J., and Toiviainen, P. (2001). Statistical features and perceived similarity of folk melodies. *Music Percept.* 18, 275–296. doi: 10.1525/mp.2001.18.3.275

Eerola, T., and Vuoskoski, J. K. (2013). A review of music and emotion studies: approaches, emotion models, and stimuli. *Music Percept.* 30, 307–340. doi: 10.1525/MP.2012.30.3.307

Eklund, R., and Mcallister, A. (2015). "An Acoustic Analys of 'Kulning' (Cattle Calls) Recorded in an Outdoor Setting on Location in Dalarna (Sweden)," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, eds M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, and J. Stuart-Smith (Glasgow: University of Glasgow), 10–14.

Eklund, R., Mcallister, A., and Dahlström, K. (2019). "An acoustic analysis of Swedish cattle calls, "kulning", performed outdoors at three distances," in *Proceedings from FONETIK 2019 Stockholm, June 10–12, 2019*, ed. M. Heldner (Stockholm: Stockholm University), 61–66.

Emsheimer, E. (ed.) (1991). "Schwedische Schellenmagie (Für Walter Wiora zum 70. Geburstag)," in *Studia Ethnomusicologica Eurasiatica*, (Stockholm: Kungliga Musikaliska Akademien), 41–52.

Etzel, J. A., Johnsen, E. L., Dickerson, J., and Adolphs, R. (2006). Cardiovascular and respiratory responses during musical mood induction. *Int. J. Psychophysiol.* 61, 57–69. doi: 10.1016/j.ijpsycho.2005.10.025

Fabian, D., and Schubert, E. (2008). Musical character and the performance and perception of dotting, articulation and tempo in 34 recordings of variation 7 from J.S. Bach's Goldberg Variations (BWV 988). *Music. Sci.* 12, 177–206. doi: 10.1177/102986490801200201

Fagg, M. C. (1997). *Rock Music*. Oxford: University of Oxford.

Fallows, D. (2001). "Metronome," in *The New Grove Dictionary of Music and Musicians*, ed. G. Grove (Oxford: Oxford University Press)

Farber, W. (1990). *Magic at the cradle: Babylonian and Assyrian lullabies*. Anthropos: Internationale zeitschrift für Völker-u Sprachenkunde.

Farmer, H. G. (1965). The éthos of antiquity. *Islamic Stud.* 4, 25–30.

Felius, M., Beerling, M.-L., Buchanan, D. S., Theunissen, B., Koolmees, P. A., and Lenstra, J. A. (2014). On the history of cattle genetic resources. *Diversity* 6, 705–750. doi: 10.3390/d6040705

Fenk-Oczlon, G., and Fenk, A. (2009a). "Musical pitch in nonsense syllables: correlations with the vowel system and evolutionary perspectives," in *Proceedings of the 7th ESCOM 2009*, eds J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, and P.-S. Eerola (Jyväskylä: University of Jyväskylä), 110–113.

Fenk-Oczlon, G., and Fenk, A. (2009b). Some parallels between language and music from a cognitive and evolutionary perspective. *Music. Sci.* 13, 201–226. doi: 10.1177/102986490901300210l

Fétis, F.-J. (1994). *Esquisse de L'histoire de L'harmonie*. Hillsdale, NY: Pendragon Press.

Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: a comparative approach to the emergence of language. *Front. Psychol.* 7:1393. doi: 10.3389/fpsyg.2016.01393

Fitch, W. T. (2006). The biology and evolution of music: a comparative perspective. *Cognition* 100, 173–215. doi: 10.1016/j.cognition.2005.11.009

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge, MA: Cambridge University Press.

Fitch, W. T. (2012). "The biology and evolution of rhythm: unravelling a paradox," in *Language and Music as Cognitive Systems*, eds P. Rebuschat, M. Rohrmeier, J. A. Hawkins, and I. Cross (Oxford: Oxford University Press), 73–95. doi: 10.1093/acprof:oso/9780199553426.003.0009

Flatz, G., and Rotthauwe, H. W. (1973). Lactose nutrition and natural selection. *Lancet* 302, 76–77. doi: 10.1016/S0140-6736(73)93267-4

Foster, M. L. (1994). "Symbolism: the foundation of culture," in *Companion Encyclopedia of Anthropology*, ed. T. Ingold (New York, NY: Routledge), 366–395.

Fraisse, P. (1982). "Rhythm and tempo," in *Psychology of Music*, ed. D. Deutsch (New York: Academic Press), 149–180. doi: 10.1016/b978-0-12-213562-0.50010-3

Franklin, J. C. (2006). Lyre gods of the bronze age musical koine. *J. Ancient Near East. Relig.* 6, 463–482.

Friberg, A., and Sundberg, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *J. Acoust. Soc. Am.* 105, 1469–1484. doi: 10.1121/1.426687

Frödin, J. (1929). *Om Fäbodbebyggelsens Utbredning Och Olika Typer i Europa*. Uppsala: Geografiska institution, 176–194.

Gabrielsson, A. (1999). "The performance of music," in *Psychology of Music*, ed. D. Deutsch (New York: Academic Press), 501–602.

Gabrielsson, A., and Juslin, P. N. (2003). "Emotional expression in music," in *Handbook of Affective Sciences Series in Affective Science*, eds R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (New York, NY: Oxford University Press), 503–534.

Gabrielsson, A., and Lindström, E. (2001). "The influence of musical structure on emotional expression," in *Music and Emotion. Theory and Research*, eds P. N. Juslin and J. A. Sloboda (Oxford, NY: Oxford University Press), 223–248.

Garbuzov, N. (1948). *Zonal Nature of Pitch Hearing [Зонная природа звуковысотного слуха]*. Moscow: Russian Academy of Science.

Garbuzov, N. (1950). *Zonal Nature of Tempo and Rhythm [Зонная природа темпа и ритма]*. Moscow: Academy of Science of the USSR.

Garbuzov, N. (1955). *Zonal Nature of Hearing of Dynamics [Зонная природа ди нами ческого слуха]*. Moscow: Gos Muz Izdat [State Musical Publishing].

Garbuzov, N. (1956). *Zonal Nature of Hearing of Timbre [Зонная природа тембрового слуха]*. Moscow: Gos Muz Izdat [State Musical Publishing].

Garroway, K. H. (2019). *Growing Up in Ancient Israel: Children in Material Culture and Biblical Texts*. Atlanta: SBL Press.

Gembris, H. (2006). "The development of musical abilities," in *MENC Handbook of Musical Cognition and Development*, ed. R. Colwell (New York, NY: Oxford University Press), 124–164. doi: 10.1093/acprof:oso/9780195304565.003.0005 doi: 10.1093/acprof:oso/9780195304565.003.0005

Gerbault, P., Roffet-Salque, M., Evershed, R. P., and Thomas, M. G. (2013). How long have adult humans been consuming milk? *IUBMB Life* 65, 983–990. doi: 10.1002/iub.1227

Germonpré, M., Láznièková-Galetová, M., and Sablin, M. V. (2012). Palaeolithic dog skulls at the gravettian pøedmostí site, the czech republic. *J. Archaeol. Sci.* 39, 184–202. doi: 10.1016/j.jas.2011.09.022

Gimbutas, M. (1993). The Indo-Europeanization of Europe: the intrusion of steppe pastoralists from south Russia and the transformation of Old Europe. *WORD* 44, 205–222. doi: 10.1080/00437956.1993.11435900

Gioia, T. (2006a). *Healing Songs*. Durham, NC: Duke University Press.

Gioia, T. (2006b). *Work Songs*. Durham, NC: Duke University Press.

Golub, H. L., and Corwin, M. J. (1985). "A physioacoustic model of the infant cry," in *Infant Crying: Theoretical and research perspectives*, eds C. F. Z. Boukydis and B. M. Lester (Boston, MA: Springer), 59–82. doi: 10.1007/978-1-4613-2381-5_3

Gordon, B. (2003). Rangifer and man: an ancient relationship. *Rangifer* 23, 15–27. doi: 10.7557/2.23.5.1651

Gourlay, K. A. (1982). Towards a humanizing ethnomusicology. *Ethnomusicology* 26, 411–420.

Gourlay, K. A. (1984). The non-universality of music and the Universality of non-music. *World Music* 26, 25–39.

Graf, W. T. (1980). *Vergleichende Musikwissenschaft: ausgewählte Aufsätze*. Vienna: Föhrenau.

Gray, R. D., and Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426, 435–439. doi: 10.1038/nature02029

Gray, R. D., Atkinson, Q. D., and Greenhill, S. J. (2011). Language evolution and human history: what a difference a date makes. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1090–1100. doi: 10.1098/rstb.2010.0378

Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2005). How does music arouse "chills"? Investigating strong emotions, combining psychological, physiological, and psychoacoustical methods. *Ann. N. Y. Acad. Sci.* 1060, 446–449. doi: 10.1196/annals.1360.041

Guilford, T., and Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Anim. Behav.* 42, 1–14. doi: 10.1016/S0003-3472(05)80600-1 doi: 10.1016/s0003-3472(05)80600-1

Halpern, A. R., Bartlett, J. C., and Dowling, W. J. (1998). Perception of mode, rhythm, and contour in unfamiliar melodies: effects of age and experience. *Music Percept.* 15, 335–355. doi: 10.2307/40300862

Hancock, A. M., Witonsky, D. B., Ehler, E., Alkorta-Aranburu, G., Beall, C., Gebremedhin, A., et al. (2010). Colloquium paper: human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proc. Natl. Acad. Sci. U.S.A.* 107(Suppl.), 8924–8930. doi: 10.1073/pnas.0914625107

Hansen, L. I., and Olsen, B. J. (2014). *Hunters in Transition: An Outline of Early Sámi History*. Leiden: Brill Publishers.

Harding, R. E. M. (1983). *The Metronome and Its Precursors*. Oxford: Gresham Books.

Hargreaves, D. J. (1996). "The development of artistic and musical competence," in *Musical Beginnings*, eds I. DelieÌge and J. Sloboda (Oxford: Oxford University Press), 145–170.

Harris, M. (1964). *The Nature of Cultural Things*. New York, NY: Random House.

Harris, M. (1990). "Emics and etics revisited," in *Emics and Etics: The Insider/Outsider Debate*, eds N. Thomas, K. Headland, L. Pike, and M. Harris (Newbury Park: SAGE Publications), 48–61.

Hauser, A. (1999). *The Social History of Art: From Prehistoric Times to the Middle Ages*. New York: Psychology Press.

Hauser, M. D. (1996). *The Evolution of Communication*. Cambridge, MA: MIT Press.

Hauser, M. D. (2000). A primate dictionary? Decoding the function and meaning of another species' vocalizations. *Cogn. Sci.* 24, 445–475. doi: 10.1016/S0364-0213(00)00026-4

Headland, T. N. (1990). "Introduction: a dialogue between Kenneth Pike and Marvin Harris on emics and etics," in *Emics and Etics: The Insider/Outsider Debate*, eds N. Thomas, K. Headland, L. Pike, and M. Harris (Newbury Park: SAGE Publications), 13–27.

Heggarty, P. (2015). "Europe and western Asia: Indo-European linguistic history," in *The Global Prehistory of Human Migration*, ed. P. S. Bellwood (Oxford UK: Wiley-Blackwell), 157–167.

Helmer, R. H. P. (1975). *European Pastoral Calls and Their Possible Influence on Western Liturgical Chant*. New York, NY: Columbia University Press.

Helskog, K. (2012). Ancient depictions of reindeer enclosures and their environment. *Fennoscandia Archaeol.* 29, 29–54.

Hershkovitz, I., Donoghue, H. D., Minnikin, D. E., May, H., Lee, O. Y.-C., Feldman, M., et al. (2015). Tuberculosis origin: the Neolithic scenario. *Tuberculosis* 95, S122–S126. doi: 10.1016/J.TUBE.2015.02.021

Hillert, D. G. (2015). On the evolving biology of language. *Front. Psychol.* 6:1796. doi: 10.3389/fpsyg.2015.01796

Hiltebeitel, A. (2011). "The indus valley 'proto-śiva': reexamined through reflections on the goddess, the buffalo, and the symbolism of våhanas," in *When the Goddess was a Woman*, eds V. Adluri and J. Bagchee (Leiden: Brill Academic Publishers), 399–431.

Hirsh-Pasek, K., and Treiman, R. (1982). Doggerel: motherese in a new context. *J. Child Lang.* 9, 229–237. doi: 10.1017/S0305000900003731

Holahan, J. M. (1987). *"Toward a Theory of Music Syntax: Some Observations of Music Babble in Young Children," in Music and Child Development*. New York, NY: Springer, 96–106.

Honing, H. (2003). The final ritard: on music, motion, and kinematic models. *Comput. Music J.* 27, 66–72. doi: 10.1162/014892603322482538

Horvaìth, T. (2012). "Animal deposits in the late copper age settlement of balatonőszöd–temetői dűlő, hungary," in *The Ritual Killing and Burial of Animals European Perspectives*, ed. A. Pluskowski (Oxford: Oxbow Books), 115–136.

Hubbard, T. L. (2003). Further correspondences and similarities of shamanism and cognitive science: mental representation, implicit processing, and cognitive structures. *Anthropol. Consciousness* 14, 40–74. doi: 10.1525/ac.2003.14.1.40

Hurford, J. R. (2012). *Language in the Light of Evolution: The Origins of Grammar*. Oxford, UK: Oxford University Press.

Huron, D. (1989). "Characterizing Musical Textures," in *Proceedings: 1989 International Computer Music Conference, November 2-5*, (Ann Arbor, MI: Michigan Publishing, University of Michigan Library), 131–134.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Ingold, T. (1986). Reindeer economies: and the origins of pastoralism. *Anthropol. Today* 2, 5–10. doi: 10.2307/3032710

Ivarsdotter, A. (1986). *Sången I Skogen. Studier Kring Den Svenska FäBodmusiken*. Dissertation, Uppsala University, Department of Musicology, Uppsala.

Ivarsdotter, A. (1995). *Lockrop & Vallåtar - Ancient Swedish Pastoral Music*. Stockholm: Musica Sveciae.

Ivarsdotter, A. (2004). "And the cattle follow her, for they know her voice . . . On communication between women and cattle in Scandinavian pastures," in *Pecus. Man and Animal in Antiquity*, ed. B. S. Frizell (Rome: The Swedish Institute in Rome), 146–149.

Jerkert, J. (2003). *Measurements and Models of Musical Articulation*. Available online at: http://www.speech.kth.se/publications/masterprojects/2004/Jerkert.pdf (accessed July 31, 2019).

Johansson, S. (2005). *Origins of Language. Constraints on Hypotheses*. Amsterdam: J. Benjamins Pub. Co.

Johnson, A. (1979). Fäbodmusik i Förvandling: Rapport Från En Exkursion Sommaren 1977. *Svensk Tidskrift Musikforskning* 60, 5–39.

Johnson, A. (1984). Voice physiology and ethnomusicology: physiological and acoustical studies of the swedish herding song. *Yearbook Tradit. Music* 16, 42–66. doi: 10.2307/768202

Johnson, A. (1990). "The sprite in the water and the siren of the woods. On Swedish folk music and gender," in *Music, Gender, and Culture*, eds M. Herndon and S. Ziegler (Wilhelmshaven: F. Noetzel), 27–40.

Johnson, A., Sundberg, J., and Wilbrand, J. (1982). Study of phonation and articulation in a type of Swedish herding song. *Dept. Speech Music Hear. Q. Prog. Status Rep.* 23, 99–116.

Johnson-Laird, P. N., and Oatley, K. (2010). "Emotions, music, and literature," in *Handbook of Emotions*, eds M. Lewis, J. M. Haviland-Jones, and L. F. Barrett (New York: The Guilford Press), 102–113.

Jones, M. R. (2016). "Musical time," in *The Oxford Handbook of Music Psychology*, eds S. Hallam, I. Cross, and M. Thaut (Oxford, UK: Oxford University Press), 125–141.

Jones, M. R., and Large, E. W. (1999). The dynamics of attending: how people track time-varying events. *Psychol. Rev.* 106, 119–159. doi: 10.1037/0033-295x.106.1.119 doi: 10.1037/0033-295x.106.1.119

Jones, M. R., and Ralston, J. T. (1991). Some influences of accent structure on melody recognition. *Mem. Cogn.* 19, 8–20. doi: 10.3758/BF03198492

Jordania, J. (2011). *Why Do People Sing: Music in Human Evolution*. Tbilisi: Logos.

Jorgensen, E. R. (2003). Western classical music and general education. *Philos. Music Educ. Rev.* 11, 130–140. doi: 10.2979/pme.2003.11.2.130

Jürgens, U. (1995). "Neuronal control of vocal production in non-human and human primates," in *Current Topics in Primate Vocal Communication*, eds E. Zimmermann, J.D. Newman, and U. Jürgens (Boston, MA: Springer), 199–206. doi: 10.1007/978-1-4757-9930-9_10

Juslin, P. N. (2001). "Communicating Emotion in Music Performance: A Review and Theoretical Framework," in *Music and Emotion: Theory and Research*, eds P. N. Juslin and J. A. Sloboda (Oxford, UK: Oxford University Press), 309–337.

Juslin, P. N. (2005). "From mimesis to catharsis: expression, perception, and induction of emotion in music," in *Musical Communication*, eds D. Miell, R. MacDonald, and D. J. Hargreaves (Oxford, UK: Oxford University Press), 85–116. doi: 10.1093/acprof:oso/9780198529361.003.0005

Juslin, P. N. (2011). "Music and emotion: seven questions, seven answers," in *Music and the Mind*, eds I. Deliège and J. Davidson (Oxford, UK: Oxford University Press), 113–138.

Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* 10, 235–266. doi: 10.1016/j.plrev.2013.05.008

Juslin, P. N., and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychol. Bull.* 129, 770–814. doi: 10.1037/0033-2909.129.5.770

Juslin, P. N., and Västfjäll, D. (2008). All emotions are not created equal: reaching beyond the traditional disputes. *Behav. Brain Sci.* 31, 600–621. doi: 10.1017/S0140525X08005554

Kaminska, Z., and Woolf, J. (2000). Melodic line and emotion: cooke's theory revisited. *Psychol. Music* 28, 133–153. doi: 10.1177/0305735600282003

Kan-ool, A. K. (2012). Pastoral plots in the traditional Tuvinian culture of the Erzya kozhuun. [С котоводческиезаговоры в традиционной купьтурету ьинцеь рзинскожууна]. *Bull. East Siberian Acad. Culture Arts* 1, 31–35.

Keller, H. (1973). *Phrasing and Articulation: A Contribution to a Rhetoric of Music, with 152 Musical Examples*. New York, NY: W.W. Norton.

Kendall, R. A., and Carterette, E. C. (1993). Identification and blend of timbres as a basis for orchestration. *Contemp. Music Rev.* 9, 51–67. doi: 10.1080/07494469300640341

Khannanov, I. I. (2005). *Russian Methodology of Musical Form and Analysis*. Dissertation, University of California, Santa Barbara, Department of Musicology, Santa Barbara, CA.

Kharlap, M. (1972). "Traditional Russian musical system and the problem of origin of music [Народно-русскаямузыкапьнаясистема и проб лемапоисхождениямузыки]," in *Early forms of art [ранниеформьIu скусстеа]*, ed. Y. Meletinskii (Moscow: Iskusstvo), 246–247.

Kholopov, Y. (1976). Mode[ад]. *Encyclopedia Musica[узыкалънаягнии клопедиа]* 3, 130–143.

Kholopov, Y. (2005). "Towards the problem of mode in Russian theoretic musicology [к проблемелада в русскомтеоретэче скоммузыкознании]," in *Harmony: Problems of Science and Methodology [Гармонии:про блемынауки и Методики]*, ed. E. Struchalina (Rostov-na-Donu: RGK (Rostov State Conservatory)), 135–157.

Kholopov, Y. (2006). *Introduction into Musical Form [теоруямушыки:* мело *дика,pumMuka,* факмука,ме-ми зм*]*. Moscow: Moscow State Tchaikovsky Conservatory.

Kholopova, V. (2002). *Theory of Music: Melos, Rhythm, Texture, Thematicism [Теории* ямузыки*: ме лодика, pumMuka, фактура, теМатиз М]*. Sankt-Petersburg: Lan.

Kirnarskaya, D., Kiyashchenko, N., Tarasova, K., and Tzypina, G. (2003). *Psychology of Musical Activities: Theory and Practice [психо логи* ямузыкаль ноу *деятельности: Теория и практика]*. Moscow: Akademiya.

Klein, R. G. (2009). *The Human Career: Human Biological and Cultural Origins*, 3rd Edn. Chicago, IL: University of Chicago Press.

Kolinsky, R., Lidji, P., Peretz, I., Besson, M., and Morais, J. (2009). Processing interactions between phonology and melody: vowels sing but consonants speak. *Cognition* 112, 1–20. doi: 10.1016/j.cognition.2009.02.014

Kolltveit, G. (2008). "Animal bells in early scandinavian soundscapes," in *Studies in Music Archaeology VI. Current Challenges and New Objectives in Music Archaeology*, eds E. Hickmann, J. Orlamünde, and R. Eichmann (Rahden: Marie Leidorf), 147–153.

Kondratyeva, N. M. (1989). "Lullabies of Teleuts, Tubalars, Kumandins, Chelkans [колыбельныетелеутов, тубаларов, кумандгнцев, члканцев]," in *Folkloric Heritage of the Mountain Altai [Фольклорноенаследие Горноио Алтая]*, ed. V. Edokov (Gorno-Altaisk: Gorno-Altaisk state institute of hystory, language and literature), 20–48.

Kondratyeva, N. M. (1996). *Pastoral Magic Spells of Telengits* [С котоводшескиезаГ оворытеленГі тов]. Dissertation, Novosibirsk State Conservatory named after Glinka, Department of Ethnomusicology, Novosibirsk.

Kondratyeva, N. M., and Kopytov, R. B. (2017). The experiment in notation of sung pastoral magic spells of Sutkhol Tuvans [опытнотированияпоищихсяскот оводческэх заговоровсутхо льских тувинцев]. *J. Musical Sci.* 3, 77–83.

Kondratyeva, N. M., and Mazepus, V. V. (1999). Modal organization of Telengit pastoral spells [Модальнаяорганизациятеленгитцкихскотоводческих заговоров]. *Lang. Indigenous Peoples Siberia* 6, 19–43.

Korom, F. J. (2000). Holy cow! the apotheosis of zebu, or why the cow is sacred in hinduism. *Asian Folklore Stud.* 59, 181–203. doi: 10.2307/1178915

Kozłowski, J. K. (2015). The origin of the Gravettian. *Q. Int.* 359–360, 3–18. doi: 10.1016/j.quaint.2014.03.025

Kreitner, K., Térey-Smith, M., Westrup, J., Holoman, D. K., Hopkins, G. W., Griffiths, P., et al. (2001). "Instrumentation and orchestration" in *The New Grove Dictionary of Music and Musicians*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. New York, NY: Oxford University Press.

Krumhansl, C. L. (2002). Music: a link between cognition and emotion. *Curr. Direct. Psychol. Sci.* 11, 45–50. doi: 10.1111/1467-8721.00165

Krushanov, A. I. (1987). *History and Culture of Chukchi: Historic-Ethnographic Sketches [История и культура чукчей: историко-этнографические очерки]*. Leningrad: Nauka.

Kubik, G. (1999). *Africa and the Blues*. Jackson, MS: University Press of Mississippi.

Kvamme, M. (1988). "Pollen analytical studies of mountain summer-farming in western Norway," in *The Cultural Landscape, Past, Present and Future*, eds H. H. Birks, H. J. B. Birks, P. E. Kaland, and D. Moe (Cambridge, MA: Cambridge University Press), 349–365.

Kyrgys, Z. K. (2002). *Tuvan Throat Singing [Туеинскоегорлоеоепение]*. Russia: Nauka.

Large, E. W. (2008). "Resonating to musical rhythm: theory and experiment," in *Psychology of Time*, ed. S. Grondin (Bingley: Emerald Group Publishing Limited), 189–231.

Large, E. W., and Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connect. Sci.* 6, 177–208. doi: 10.1080/09540099408915723

Larson, G., Karlsson, E. K., Perri, A., Webster, M. T., Ho, S. Y. W., Peters, J., et al. (2012). Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proc. Natl. Acad. Sci. U.S.A.* 109, 8878–8883. doi: 10.1073/pnas.1203005109

Lerdahl, F., and Jackendoff, R. S. (1985). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Lester, J. (1989). *Between Modes and Keys: German Theory, 1592-1802*. New York, NY: Pendragon Press.

Leuthold, H. (1981). *Der Naturjodel in der Schweiz: Wesen, Entstehung, Charakteristik, Verbreitung: ein Forschungsergebnis über den Naturjodel in der Schweiz*. Altdorf: Robert Fellmann Liederverlag.

Levin, T. C., and Suzukei, V. (2006). *Where Rivers and Mountains Sing: Sound, Music, and Nomadism in Tuva and Beyond*. Bloomington, IN: Indiana University Press.

Levitin, D. J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Percept. Psychophys.* 56, 414–423. doi: 10.3758/BF03206733

Levitin, D. J. (2009). *The World in Six Songs: How the Musical Brain Created Human Nature*. New York, NY: Plume Books.

Lieberman, P. (1985). "The physiology of cry and speech in relation to linguistic behavior," in *Infant Crying: Theoretical and Research Perspectives*, eds C. F. Z. Boukydis and B. Lester (Boston, MA: Springer), 29–57. doi: 10.1007/978-1-4613-2381-5_2

Lodrick, D. O. (2005). Symbol and sustenance: cattle in South Asian culture. *Dialect. Anthropol.* 29, 61–84. doi: 10.1007/s10624-005-5809-8

Loewy, J. V. (1995). The musical stages of speech: a developmental model of pre-verbal sound making. *Music Ther.* 13, 47–73. doi: 10.1093/mt/13.1.47

Lomax, A. (1977). Universals in Song. *World Music* 19, 117–129.

London, J. (2004). *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford, New York: Oxford University Press.

Lõugas, L., Kriiska, A., and Maldre, L. (2007). New dates for the Late Neolithic Corded Ware Culture burials and early husbandry in the East Baltic region. *Archaeofauna* 16, 21–31.

Loui, P., Patterson, S., Sachs, M. E., Leung, Y., Zeng, T., and Przysinda, E. (2017). White matter correlates of musical Anhedonia: implications for evolution of music. *Front. Psychol.* 8:1664. doi: 10.3389/fpsyg.2017.01664

Madison, G., and Paulin, J. (2010). Ratings of speed in real music as a function of both original and manipulated beat tempo. *J. Acoust. Soc. Am.* 128, 3032–3040. doi: 10.1121/1.3493462

Maier, A. (2017). Population and settlement dynamics from the gravettian to the magdalenian. *Mitteilungen Gesellschaft Urgeschichte* 26, 83–102.

Malmström, H., Linderholm, A., Lidén, K., Storå, J., Molnar, P., Holmlund, G., et al. (2010). High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. *BMC Evol. Biol.* 10:89. doi: 10.1186/1471-2148-10-89

Manser, M. B. (2010). "The generation of functionally referential and motivational vocal signals in mammals," in *Handbook of Behavioral Neuroscience*, ed. S. M. Brudzynski (New York, NY: Elsevier), 477–486. doi: 10.1016/b978-0-12-374593-4.00043-7

Manuel, P., and Blum, S. (2011). "Classical aesthetic traditions of India, China, and the Middle East," in *The Routledge Companion to Philosophy and Music*, eds T. Gracyk and A. Kania (Abingdon: Routledge), 245–256.

Marciniak, A. (2013). "Origin of stock-keeping and the spread of animal exploitation strategies in the early and middle Neolithic of the North European Plain," in *Origins and Spread of Domestic Animals in Southwest Asia and Europe*, eds S. Colledge, J. Conolly, K. Dobney, and S. Shennan (Walnut Creek, CA: Left Coast Press), 221–236.

Marciniak, S., and Perry, G. H. (2017). Harnessing ancient genomes to study the history of human adaptation. *Nat. Rev. Genet.* 18, 659–674. doi: 10.1038/nrg.2017.65

Marler, P. (1997). Three models of song learning: evidence from behavior. *J. Neurobiol.* 33, 501–516. doi: 10.1002/(sici)1097-4695(19971105)33:5⟨501::aid-neu2⟩3.0.co;2-8

Marler, P. (2001). "Origins of music and speech: insights from animals," in *The Origins of Music*, eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 31–48.

Marler, P., and Mitani, J. C. (2008). A phonological analysis of male gibbon singing behavior. *Behaviour* 109, 20–45. doi: 10.1163/156853989x00141

Marler, P., and Slabbekoorn, H. W. (2004). *Nature's Music: the Science of Birdsong*. Amsterdam: Elsevier.

Marlowe, F. W. (2005). Hunter-gatherers and human evolution. *Evol. Anthropol. Issues News Rev.* 14, 54–67. doi: 10.1002/evan.20046

Marshack, A. (1990). "Early hominid symbol and evolution of the human capacity," in *The Emergence of Modern Human. An Archaeological Perpective*, ed. P. Mellars (Edinburgh: Edinburgh University Press), 457–498.

Mathiesen, T. J. (1984). Harmonia and ethos in ancient Greek music. *J. Musicol.* 3, 264–279. doi: 10.1525/jm.1984.3.3.03a00050

Mathiesen, T. J. (1999). *Apollo's Lyre: Greek Music and Music Theory in Antiquity and the Middle Ages*. Lincoln, NE: University of Nebraska Press.

Mattheson, J., and Harriss, E. C. (1981). *Johann Mattheson's Der Vollkommene Capellmeister: A Revised Translation With Critical Commentary*. Ann Arbor, Mich.: UMI Research Press.

Maynard-Smith, J. (1976). Evolution and the theory of games. *Am. Sci.* 64, 41–45.

Maynard-Smith, J., and Harper, D. (2003). *Animal Signals*. Oxford: Oxford University Press.

Mazel, L. (1979). *Structuring of Musical Works [Строениету звлка льп ыхпроизеедении]*, 2nd Edn. Moscow: Muzyka.

Mazel, L., and Tzukkerman, V. (1967). *Analysis of Musical Works [Анализмуз влкальныхпроизеедении]*. Moscow: Muzyka.

Mazepus, V. V. (1993). *Universal Grammatic Approach to Culturology* [Универсально-грамматический подход в культурологии] . Novosibirsk: State Conservatory named after Glinka.

Mazepus, V. V., and Galitskaya, S. P. (1997). *Musical Culture of Siberia. Traditional Culture of Indigenous People of Siberia [Музвлкальнаякульмураси бири. традииионнаякульмуракоренныхНародоеСи бири]*. Novosibirsk: Novosibirsk State Conservatory named after Glinka.

McAdams, S. (2004). Perception of musical similarity among contemporary thematic materials in two instrumentations. *Music Percept. Interdiscip. J.* 22, 207–237. doi: 10.1525/mp.2004.22.2.207

McAuley, J. D. (2010). "Tempo and rhythm," in *Springer Handbook of Auditory Research*, eds J. M. Riess, R. Fay, and A. Popper (New York: Springer), 165–199.

Mcbrearty, S., and Brooks, A. S. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J. Hum. Evol.* 39, 543–563. doi: 10.1006/JHEV.2000.0435

McConnell, P. B. (1990). Acoustic structure and receiver response in domestic dogs, Canis familiaris. *Anim. Behav.* 39, 897–904. doi: 10.1016/S0003-3472(05)80954-6

McConnell, P. B. (1991). Lessons from animal trainers: the effect of acoustic structure on an animal's response. *Perspect. Ethol.* 9, 165–187.

McConnell, P. B. (2002). *The Other End of the Leash: Why We Do What We Do Around Dogs*. New York: Ballantine Books.

McConnell, P. B., and Baylis, J. R. (1985). Interspecific communication in cooperative herding: acoustic and visual signals from human shepherds and herding dogs. *Zeitschrift Tierpsychol.* 67, 302–328. doi: 10.1111/j.1439-0310.1985.tb01396.x

McCreless, P. (2002). "Music and rhetoric," in *The Cambridge History of Western Music Theory* (Cambridge: Cambridge University Press), 847–879. doi: 10.1111/j.1439-0310.1985.tb01396.x

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Curr. Biol. Biol.* 20, 1035–1041. doi: 10.1016/j.cub.2010.04.019

McInerney, J. (2010). *The Cattle of the Sun: Cows and Culture in the World of the Ancient Greeks*. Princeton, NJ: Princeton University Press.

Medushevsky, V. V. (2010). *On Rules and Means of Artistic Influence of Music [О закономерностях и средствах художественного воздействия музыки]*. Moscow: Muzyka.

Meier, B., and Dittmer, L. A. (1956). The musica reservata of adrianus petit coclico and its relationship to josquin. *Musica Discip.* 10, 67–105.

Mellars, P. (2004). "Archaeology and the origins of modern humans: European and African perspectives," in *The Speciation of Modern Homo Sapiens*, ed. T. J. Crow (London: British Academy), 31–47.

Mertens, P. (2004). "The prosogram: semi-automatic transcription of prosody based on a tonal perception model," in *Proceedings of the 2nd International Conference on Speech Prosody*, Nara, 549–552.

Messner, G. F. (1993). Ethnomusicological research, another "Performance" in the international year of indigenous peoples? *World Music* 35, 81–95.

Meyer, J. (2009). *Acoustics and The Performance of Music Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instruments Makers. 5th (1978)*. Berlin: Springer.

Miklosi, A. (2015). *Dog Behaviour, Evolution, and Cognition*. Oxford: Oxford University Press.

Miller, R. J. (2014). *Contemporary Orchestration: A Practical Guide to Instruments, Ensembles, and Musicians*. New York, NY: Routledge.

Mills, S. (2016). *Auditory Archaeology*. London: Routledge.

Milne, L. S. (2017). The terrors of the night: charms against the nightmare and the mythology of dreams. incantatio. *Int. J. Charms Charm. Charm.* 6, 78–116. doi: 10.7592/incantatio2017_6_milne

Mitchell, R. W. (2001). Americans' talk to dogs: similarities and differences with talk to infants. *Res. Lang. Soc. Interact.* 34, 183–210. doi: 10.1207/S15327973RLSI34-2_2

Mitchell, S. A. (2001). Warlocks, valkyries and varlets: a prolegomenon to the Study of North Sea witchcraft terminology. *Cosmos* 17, 59–81.

Moberg, C.-A. (1955). Om vallåtar. En studie i de svenska fäbodarnas musikaliska organisation [On herding calls: a study of the musical organization of the Swedish summe]. *Svensk Tidskrift r Musikforskning* 37, 1–27.

Moberg, C.-A. (1971). *Studien zur schwedischen Volksmusik*. Uppsala: Almquist & Wiksell.

Mohn, C., Argstatter, H., and Wilker, F.-W. (2010). Perception of six basic emotions in music. *Psychol. Music* 39, 503–517. doi: 10.1177/0305735610378183

Molino, J. (1990). Musical fact and the semiology of music. *Music Anal.* 9:105. doi: 10.2307/854225

Monahan, C. B., and Carterette, E. C. (1985). Pitch and duration as determinants of musical space. *Music Percept. Interdiscip. J.* 3, 1–32. doi: 10.2307/40285320 doi: 10.2307/40285320

Monahan, C. B., Kendall, R. A., and Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Percept. Psychophys.* 41, 576–600. doi: 10.3758/BF03210491

Monelle, R. (1992). *Linguistics and Semiotics in Music*. Reading: Harwood Academic Publishers.

Monelle, R. (2000). *The Sense of Music: Semiotic Essays*. Princeton, NJ: Princeton University Press.

Monelle, R. (2006). *The Musical Topic: Hunt, Military and Pastoral*. Indianapolis, IN: Indiana University Press.

Montagu, J. (2004). How old is music? *Galpin Soc. J.* 57, 171–182.

Moog, H. (1976). *The Musical Experience of the Pre-School Child*. London: Schott Music.

Moore, A. F. (2011). The end of the revival: the folk aesthetic and its 'mutation.'. *Popular Music History* 4, 289–307. doi: 10.1558/pomh.v4i3.289

Morley, I. (2006). Mousterian musicianship? The case of the divje babe i bone. *Oxf. J. Archaeol.* 25, 317–333. doi: 10.1111/j.1468-0092.2006.00264.x

Morley, I. (2013). *The Prehistory of Music: Human Evolution, Archaeology, and the Origins of Musicality*. Oxford: Oxford University Press.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *Am. Nat.* 111, 855–869. doi: 10.1086/283219

Münzel, S., and Conard, N. J. (2009). "Flötenklang aus fernen zeiten die frühesten musikinstrumente," in *Eiszeit. Kunst und Kultur. Begleitband zur großen Landesausstellung*, ed. E. K. U. Tübingen (Ostfildern: Archäologisches Landesmuseum Baden Württemberg), 317–321.

Murashkin, A. I., Kolpakov, E. M., Shumkin, V. Y., Khartanovich, V. I., and Moiseyev, V. G. (2016). Kola oleneostrovskiy grave field: a unique burial site in the european arctic. *ISKOS* 21, 185–199.

Myers, H. (1993). *Ethnomusicology: Historical and Regional Studies*. New York, NY: Norton.

Naguib, M., and Riebel, K. (2014). "Singing in space and time: the biology of birdsong," in *Biocommunication of Animals*, ed. G. Witzany (Dordrecht: Springer), 233–247. doi: 10.1007/978-94-007-7414-8_13

Nakata, T., and Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behav. Dev.* 27, 455–464. doi: 10.1016/j.infbeh.2004.03.002 doi: 10.1016/j.infbeh.2004.03.002

Narmour, E. (1992). *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago, IL: Chicago Review Press.

Nash, P. H., and Carney, G. O. (1996). The seven themes of music geography. *The Canadian Geographer/Le Gographe Canadien* 40, 69–74. doi: 10.1111/j.1541-0064.1996.tb00433.x

Nattiez, J.-J. (1990). *Music and Discourse: Toward a Semiology of Music*. Princeton, NJ: Princeton University Press.

Nattiez, J.-J. (2012). Is the search for universals incompatible with the study of cultural specificity? *Hum. Soc. Stud.* 1, 67–94. doi: 10.2478/v10317-012-0005-2

Nawrot, E. S. (2003). The perception of emotional expression in music: evidence from infants, children and adults. *Psychol. Music* 31, 75–92. doi: 10.1177/0305735603031001325

Nazaikinsky, Y. V. (1972). *On Psychology of Human Musical Perception [О психологии музыкальноэо восприятия]*. Moscow: Muzyka.

Nazaikinsky, Y. V. (1973). "On constants in perception of music [О Концтантности в Восприятии Музыки]," in *Musical Art and Science [Музыкальное искусство и наука]*, ed. Y. V. Nazaikinsky (Moscow: Muzyka), 59–98.

Nazaikinsky, Y. V. (1982). *The Logic of Musical Composition [ Логи ка музыкальной композиции]*. Moscow: Muzyka.

Nazaikinsky, Y. V. (1988). *The Sonic World of Music [ Звуковой мир музыки]*. Moscow: Muzyka.

Nazaikinsky, Y. V. (2013). *Style and Genre in Music [ Смиль и эсанр в музыке]*. Tbilisi: Tbilisi State Conservatoire.

Nazaikinsky, Y. V., and Rags, Y. N. (1964). "Perception of musical timbres and the significance of the individual harmonics in a sound [восприятиемузыкальныхтембров изначениеотдельныхгармоник звука]," in *Application of the acoustic methods in musicology* [ *Применени еакусмически хмемодов в музыкознании*], ed. S. S. Skrebkov (Moscow: Muzyka), 79–100.

Nettl, B. (2005). *The Study of Ethnomusicology: Thirty-one Issues and Concepts*. Champaign, IL: University of Illinois Press.

Nettl, B. (2010). *Nettl's Elephant: On the History of Ethnomusicology*. Champaign, IL: University of Illinois Press.

Nielsen, S. (1997). "Worksongs in Denmark," in *Historical studies on folk and traditional music: ICTM Study Group on Historical Sources of Folk Music, conference report, Copenhagen, 24-28 April 1995*, eds D. Stockmann and J. H. Koudal (Copenhagen: Danish Folklore Archives), 241–247.

Nikolsky, A. (2015a). "эCómo funciona la emoción musical? [How Emotion Can Be the Meaning of a Music Work]," in *Música y cuerpo: Estudios musicológicos*, ed. T. Cascudo (Baleares: Calanda Ediciones Musicales), 241–262.

Nikolsky, A. (2015b). Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. Part-1: prehistoric. *Front. Psychol.* 6:1405. doi: 10.3389/fpsyg.2015.01405

Nikolsky, A. (2016). Evolution of tonal organization in music optimizes neural mechanisms in symbolic encoding of perceptual reality. part-2: ancient to seventeenth century. *Front. Psychol.* 7:211. doi: 10.3389/fpsyg.2016.00211

Nikolsky, A. (2018). Commentary: the 'Musilanguage' model of language evolution. *Front. Psychol.* 9:75. doi: 10.3389/fpsyg.2018.00075

Nikolsky, A. (2020). "Emergence of the distinction between 'verbal' and 'musical' in early childhood development," in *The Origins of Language Revisited: Differentiation from Music and the Emergence of Neurodiversity and Autism*, ed. N. Masataka (Singapore: Springer).

Nikolsky, A., Alekseyev, E. Y., Alekseev, I. Y., and Dyakonova, V. E. (2020). The overlooked tradition of 'personal music' and its place in the evolution of music. *Front. Psychol.* 10:3051. doi: 10.3389/fpsyg.2019.03051

Noorden, L. (1975). *Temporal Coherence in the Perception of Tone Sequences*. Eindhoven: Institute for Perceptual Research.

Novik, Y. S. (1999). "Semiotic functions of human voice in folklore and beliefs of Siberian peoples [Семиотическиефункцииголоса в фольклоре и верованияхнародовСибири]," in *Eastern Folklore and Mythology from the Comparative-Typological Aspect*, eds N. Lidova and N. Nikulin (Moscow: Russian Academy of Science), 217–235.

Novik, Y. S. (2004). *Rite and Folklore in Siberian Shamanism: An Experiment in Correlation of Structures [ Обряд и фольклор в си би рскомэаманизме: Оп ымсоп осмавлени ясмрукмур]*. Moscow: Eastern Literature, Russian Academy of Science.

Owren, M. J., and Rendall, D. (2001). Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evol. Anthropol. Issues News Rev.* 10, 58–71. doi: 10.1002/evan.1014 doi: 10.1002/evan.1014

Oxenham, A. J. (2013). "The perception of musical tones," in *The Psychology of Music*, ed. D. Deutsch (San Diego, CA: Academic Press), 1–33. doi: 10.1016/b978-0-12-381460-9.00001-8

Panksepp, J., and Bernatzky, G. (2002). Emotional sounds and the brain: the neuro-affective foundations of musical appreciation. *Behav. Process.* 60, 133–155. doi: 10.1016/S0376-6357(02)00080-3

Panteleeva, O. (2019). How soviet musicology became marxist. *Slavonic East Eur. Rev.* 97:73. doi: 10.5699/slaveasteurorev2.97.1.0073

Papoušek, H., and Papoušek, M. (1995). "Beginning of Human Musicality," in *Music and The Mind Machine: The Psychophysiology and Psychopathology of the Sense of Music*, ed. R. Steinberg (Berlin: Springer), 27–34. doi: 10.1007/978-3-642-79327-1_3

Parncutt, R., and Hair, G. (2011). Consonance and dissonance in music theory and psychology: disentangling dissonant dichotomies. *J. Interdiscip. Music Stud.* 5, 119–168.

Patel, A. D. (2006). An empirical method for comparing pitch patterns in spoken and musical melodies: a comment on Pearl's "Eavesdropping with a master: leos Janácek and the music of speech". *Empiric. Musicol. Rev.* 1, 166–169. doi: 10.18061/1811/24011

Patel, A. D. (2008). "Music as a transformative technology of the mind," in *Music: Its Evolution, Cognitive Basis, and Spiritual Dimensions*, ed. M. A. Meyers (Cambridge, MA: John Templeton Foundation).

Patel, A. D. (2010). "Music, biological evolution, and the brain," in *Emerging Disciplines: Shaping New Fields of Scholarly Inquiry in and Beyond the Humanities*, eds C. Levander and C. Henry (Houston, TX: Rice University Press), 91–144.

Patterson, R. D., Gaudrain, E., and Walters, T. C. (2010). "The perception of family and register in musical tones," in *Springer Handbook of Auditory Research*, eds M. Riess Jones, R. Fay, and A. Popper (New York, NY: Springer), 13–50. doi: 10.1007/978-1-4419-6114-3_2

Payne, K. (2001). "The progressively changing songs of humpback whales: a window on the creative process in a wild animal," in *The Origins of Music*, eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 135–150.

Pegg, C. (1992). Mongolian conceptualizations of overtone singing (xöömii). *Br. J. Ethnomusicol.* 1, 31–54. doi: 10.1080/09681229208567199

Pegg, C. (2001). *Mongolian Music, Dance and Oral Narrative: Performing Diverse Identities*. Seattle, WA: University of Washington Press.

Peretz, I. (2013). "Towards a neurobiology of musical emotions," in *Handbook of Music and Emotion: Theory, Research, Applications*, eds P. N. Juslin and J. A. Sloboda (Oxford: OUP Oxford), 99–126. doi: 10.1093/acprof:oso/9780199230143.003.0005

Perlovsky, L. (2012). Cognitive function, origin, and evolution of musical emotions. *Music. Sci.* 16, 185–199. doi: 10.1177/1029864912448327

Peters, G. (1984). On the structure of friendly close range vocalizations in terrestrial carnivores (Mammalia: Carnivora: Fissipedia). *Zeitschrift Säugetierkunde* 49, 157–182.

Peters, J., Pöllath, N., and Arbuckle, B. (2017). "The emergence of livestock husbandry in Early Neolithic Anatolia," in *The Oxford Handbook of Zooarchaeology*, eds U. Albarella, M. Rizzetto, H. Russ, K. Vickers, and S. Viner-Daniels (Oxford, UK: Oxford University Press), 247–265.

Pettitt, P. (2010). *The Palaeolithic Origins Of Human Burial*. London: Routledge.

Pfordresher, P. Q., and Brown, S. (2017). Vocal mistuning reveals the origin of musical scales. *J. Cogn. Psychol.* 29, 35–52. doi: 10.1080/20445911.2015.1132024 doi: 10.1080/20445911.2015.1132024

Pike, K. L. (1967). *Language in Relation to a Unified Theory of the Structure of Human Behavior*. The Hague: Mouton de Gruyter.

Pike, K. L. (1990). "On the emics and etics of Pike and Harris," in *Emics and Etics: The Insider/Outsider Debate*, eds N. Thomas, K. Headland, L. Pike, and M. Harris (Newbury Park: SAGE Publications), 28–47.

Pikovsky, A., Rosenblum, M., and Kurths, J. (2001). *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge, MA: Cambridge University Press.

Plantenga, B. (2004). *Yodel-Ay-Ee-Oooo. The secret History of yodeling around the World*. Abingdon: Routledge.

Plantinga, T. S., Alonso, S., Izagirre, N., Hervella, M., Fregel, R., van der Meer, J. W., et al. (2012). Low prevalence of lactase persistence in Neolithic South-West Europe. *Eur. J. Hum. Genet.* 20, 778–782. doi: 10.1038/ejhg.2011.254

Plotnikova, A. A. (1999a). "On the symbolism of whistling [О символикесвиста]," in *The World of Sound and Silence: Semiotics of Soound and Speech in Traditional Culture of Slavs [Мирзвучащий и молчащий: Семиомиказвука и речив мради ционнойкульмуреславян]*, ed. S. M. Tolstaya (Moscow: Indrik), 295–304.

Plotnikova, A. A. (1999b). "The sonic magic in Slavic pastoral rituals [Магиязвука вславянскойскотоводческойобрядности]," in *The World of Sound And Silence: Semiotics of Soound and Speech in Traditional Culture of Slavs [Мирзвучащий и Молчащий: Семиомиказвука и речи в мрад и ци онной кульмуре славян]*, ed. S. M. Tolstaya (Moscow: Indrik), 73–84.

Porter, J., Powers, H. S., Cowdery, J., Widdess, R., Davis, R., Perlman, M., et al. (2001). "Mode," in *Modal scales and traditional music. Middle East and Asia*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).

Potter, D. D., Fenwick, M., Abecasis, D., and Brochard, R. (2009). Perceiving rhythm where none exists: event-related potential (ERP) correlates of subjective accenting. *Cortex* 45, 103–109. doi: 10.1016/J.CORTEX.2008.01.004

Powell, A., Shennan, S., and Thomas, M. G. (2009). Late Pleistocene demography and the appearance of modern human behavior. *Science* 324, 1298–1301. doi: 10.1126/science.1170165

Powers, H. S., and Wiering, F. (2001). *Mode. The Term. Medieval Modal Theory. Modal Theories and Polyphonic Music*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).

Prakash, O. (1961). *Food and Drinks in Ancient India (From Earliest Times to 200 A. D.)*. Delhi: Munshi Ram Manohar Lal.

Prince, J. B. (2014). Contributions of pitch contour, tonality, rhythm, and meter to melodic similarity. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 2319–2337. doi: 10.1037/a0038010

Prince, J. B., Schmuckler, M. A., and Thompson, W. F. (2009). The effect of task and pitch structure on pitch-time interactions in music. *Mem. Cogn.* 37, 368–381. doi: 10.3758/MC.37.3.368

Probst, E. (1991). *Deutschland in der Steinzeit: Jäger, Fischer und Bauern zwischen Nordseeküste und Alpenraum*. Munich: Carl Bertelsmann.

Quam, R. M., Martínez, I., Rosa, M., and Arsuaga, J. L. (2017). "Evolution of hearing and language in fossil hominins," in *Primate Hearing and Communication*, eds R. M. Quam, M. Rosa, and J. L. Arsuaga (Berlin: Springer), 201–231. doi: 10.1007/978-3-319-59478-1_8

Rags, Y. N. (1980). *Garbuzov N.A. - Musician, researcher and pedagogue [Гарбузов Н.А.-Музыкант, исследователь, педагог]*. Moscow: Muzyka.

Rags, Y. N. (1999). *Aesthetics from the bottom and aesthetics from the top - quantitative ways of approaching* [Эстетика снизу и эстетика сверху – квантитативные пути сближения]. Moscow: Nauchnyi mir.

Ratner, L. G. (1980). *Classic Music: Expression, form, and Style*. New York, NY: Schirmer Books.

Ravens, S. (2014). *The Supernatural Voice: A History of High Male Singing*. Woodbridge: The Boydell Press.

Ray, K., and Thomas, J. (2003). "In the kinship of cows: the social centrality of cattle in the earlier Neolithic of southern Britain," in *Food, Culture and Identity in the Neolithic and Early Bronze Age*, ed. M. Parker-Pearson (Oxford: Hadrian Books), 37–52.

Reimers, E., and Colman, J. E. (2009). Reindeer and caribou (*Rangifer tarandus*) response towards human activities. *Rangifer* 26, 55–71. doi: 10.7557/2.26.2.188

Renfrew, C. (1987). *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Jonathan Cape.

Repp, B. H. (1995). Acoustics, perception, and production of legato articulation on a digital piano. *J. Acoust. Soc. Am.* 97, 3862–3874. doi: 10.1121/1.413065

Repp, B. H. (1998). *Perception and Production of Staccato articulation on the Piano*. New Haven, CT: Haskins Laboratories.

Réti, R. (1951). *The Thematic Process in Music*. New York, NY: Macmillan Publishers.

Reznikoff, I. (2008). Sound resonance in prehistoric times: a study of Paleolithic painted caves and rocks. *J. Acoust. Soc. Am.* 123, 3603–3603. doi: 10.1121/1.2934773

Rice, T. (1987). Toward the remodeling of ethnomusicology. *Ethnomusicology* 31, 469–488. doi: 10.2307/851667

Richerson, P. J., and Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago, IL: University of Chicago Press.

Richerson, P. J., Boyd, R., and Bettinger, R. L. (2009). Cultural innovations and demographic change. *Hum. Biol.* 81, 211–235. doi: 10.3378/027.081.0306

Robb, J. (1991). Random causes with directed effects: the Indo-European language spread and the stochastic loss of lineages. *Antiquity* 65, 287–291. doi: 10.1017/S0003598X00079758

Robb, J. (1993). A social prehistory of European languages. *Antiquity* 67, 747–760. doi: 10.1017/S0003598X00063766

Røed, K. H., Bjørklund, I., and Olsen, B. J. (2018). From wild to domestic reindeer – Genetic evidence of a non-native origin of reindeer pastoralism in northern Fennoscandia. *J. Archaeol. Sci. Rep.* 19, 279–286. doi: 10.1016/j.jasrep.2018.doi: 10.1016/j.jasrep.2018.02.04802.048

Røed, K. H., Flagstad, Ø, Nieminen, M., Holand, Ø, Dwyer, M. J., Røv, N., et al. (2008). Genetic analyses reveal independent domestication origins of Eurasian reindeer. *Proc. R. Soc. B Biol. Sci.* 275, 1849–1855. doi: 10.1098/rspb.2008.0332 doi: 10.1098/rspb.2008.0332

Roesner, E. H. (2001). *Rhythmic Modes (Modal Rhythm). The New Grove Dictionary of Music and Musicians*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).

Rosenberg, S. (2003). *Kulning. Musiken och metoden*. Stockholm: Udda Toner.

Rosenberg, S. (2014). Kulning – an ornamentation of the surrounding emptiness: about the unique Scandinavian herding calls. *Voice Speech Rev.* 8, 100–105. doi: 10.1080/23268263.2013.829712

Rothstein, W. N. (1989). *Phrase Rhythm in Tonal Music*. New York, NY: Schirmer Books.

Rouget, G. (2011). Musical efficacy: musicking to survive-the case of the pygmies. *Yearbook Tradit. Music* 43, 89–121. doi: 10.5921/yeartradmusi.43.0089

Rowley-Conwy, P. (2011). Westward Ho! The spread of agriculture from central Europe to the Atlantic. *Curr. Anthropol.* 52, S431–S451. doi: 10.1086/658368

Rowley-Conwy, P. (2013). "North of the frontier: early domestic animals in northern Europe," in *Origins and Spread of Domestic Animals in Southwest Asia and Europe*, eds S. Colledge, J. Conolly, K. Dobney, K. Manning, and S. Shennan (Walnut Creek, CA: Left Coast Press), 283–312.

Ruwet, N., and Everist, M. (1987). Methods of analysis in musicology. *Music Anal.* 6, 11–36. doi: 10.2307/854214

Sachs, M. E., Damasio, A. R., and Habibi, A. (2015). The pleasures of sad music: a systematic review. *Front. Hum. Neurosci.* 9:404. doi: 10.3389/fnhum.2015.00404

Sadie, S. (2001). "Movement," in *The New Grove Dictionary of Music and Musicians, Stanley*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).

Samson, J. (2001). *Genre. The New Grove Dictionary of Music and Musicians*, eds S. Sadie and J. Tyrrel (London: Macmillan Publishers).

Saña, M. (2013). "Domestication of animals in the Iberian Peninsula," in *Origins and Spread of Domestic Animals in Southwest Asia and Europe*, eds S. Colledge, J. Conolly, K. Dobney, K. Manning, and S. Shennan (Walnut Creek, CA: Left Coast Press), 195–220.

Sandell, G. J. (1995). Roles for spectral centroid and other factors in determining "Blended" instrument pairings in orchestration. *Music Percept. Interdiscip. J.* 13, 209–246. doi: 10.2307/40285694

Savage, P. E., and Brown, S. (2013). Toward a new comparative musicology. *Anal. Approach. World Music* 2, 148–197.

Schiavio, A., van der Schyff, D., Cespedes-Guevara, J., and Reybrouck, M. (2016). Enacting musical emotions, sense-making, dynamic systems, and the embodied mind. *Phenomenol. Cogn. Sci.* 16, 785–809. doi: 10.1007/s11097-016-9477-8

Schubert, E., and McPherson, G. E. (2015). "Underlying mechanisms and processes in the development of emotion perception in music," in *The Child as Musician: A handbook of musical development*, ed. G. E. McPherson (Oxford: Oxford University Press), 221–243. doi: 10.1093/acprof:oso/9780198744443.003.0012

Schulting, R. (2013). "On the northwestern fringes: earlier Neolithic subsistence in Britain and Ireland as seen through faunal remains and stable isotopes," in *Origins and Spread of Domestic Animals in Southwest Asia and Europe*, eds S. Colledge, J. Conolly, K. Dobney, K. Manning, and S. Shennan (Walnut Creek, CA: Left Coast Press), 313–338.

Searcy, W. A. (1992). "Measuring responses of female birds to male song," in *Playback and Studies of Animal Communication*, ed. P. K. McGregor (Boston, MA: Springer), 175–189. doi: 10.1007/978-1-4757-6203-7_12

Sebeok, T. (1994). *Signs: An Introduction to Semiotics*. Toronto: University of Toronto Press.

Seyfarth, R. M., and Cheney, D. L. (2017). Precursors to language: social cognition and pragmatic inference in primates. *Psychon. Bull. Rev.* 24, 79–84. doi: 10.3758/s13423-016-1059-9

Shatkovsky, G. I. (1986). *The Development of Musical Hearing and Skills [Развитие музыкального слуха и навыков]*. Moscow: Russian Ministry of Culture.

Sheikin, Y. I. (2002). *History of Music Culture of Siberian Ethnicities: A Comparative Historic Investigation [История музыкальной культуры народов Сибири: сравнительно-историческоеисследоваие]*. Moscow: Russian Academy of Science.

Sheikin, Y. I. (2017). *History of World Musical Culture: From Antiquity to 16th Century AD. [История мировой музыкальной культуры: от древних времен до XVI века н.э.]*. Yakutsk: The Ministry of culture and spiritual development of Sakha Republic.

Shestakov, V. P. (1975). *From Ethos to Affect. History of musical aesthetics from Antiquity to the 18th century [От гтоса к Аффекту. История музыкальной гстетики от античности до XVIII века]*. Moscow: Muzyka.

Shevtsov, V. N. (1988). "Hunting and pastoral sound imitations and cries among the Khakas [Охотничье-скотоводчекие звукоподражания и возгласы у хакасов]," in *Musical Ethnography of Northern Asia [Музыкальная этноэрафия Северной Азии]*, ed. Y. I. Sheikin (Novosibirsk: Novosibirsk State Conservatory named after Glinka), 108–129.

Shvachkin, N. H. (1948). "Development of phonematic perception in early childhood [Развитие фонематического восприятия речи в раннем возрасте]," in *The Problems of Psychology of Perception and Cognition. Works of the Institute of Psychology [Вопросы Психологии Восприятия и Мышления. Труды Института Психолоэии.]*, ed. B. Teplov (Moscow: Izvestiya of Academy of Pedagogical Sciences of Russia), 101–133.

Sicoli, M. A. (2015). "Voice registers," in *The Handbook of Discourse Analysis*, eds D. Tannen, H. E. Hamilton, and D. Schiffrin (Chichester, MA: John Wiley & Sons), 105–126. doi: 10.1002/9781118584194.ch5

Sikora, M., Pitulko, V. V., Sousa, V. C., Allentoft, M. E., Vinner, L., Rasmussen, S., et al. (2019). The population history of northeastern Siberia since the Pleistocene. *Nature* 570, 182–188. doi: 10.1038/s41586-019-1279-z

Silvia, P. J., and Nusbaum, E. C. (2011). On personality and piloerection: individual differences in aesthetic chills and other unusual aesthetic experiences. *Psychol. Aesth. Creat. Arts* 5, 208–214. doi: 10.1037/a0021914

Simpson, B. S. (1997). Canine communication. The veterinary clinics of North America. *Small Anim. Pract.* 27, 445–464. doi: 10.1016/S0195-5616(97)50048-9

Sjögren, K.-G., and Price, T. D. (2013). A complex Neolithic economy: isotope evidence for the circulation of cattle and sheep in the TRB of western Sweden. *J. Archaeol. Sci.* 40, 690–704. doi: 10.1016/j.jas.2012.08.001

Skrebkova-Filatova, M. S. (1985). *Texture in music [Фактура в музыке]*. Moscow: Muzyka.

Skrede, A. M. (2005). "Shielings and landscape in western Norway-Research traditions and recent trends," in *"Utmark" - the Outfield as Industry and Ideology in the Iron Age and the Middle Ages*, eds K. A. Bergsvik and A. Engevik (Bergen: Universitetet i Bergen), 31–41.

Slater, P. (2001). "Birdsong repertoires: their origin and use," in *The Origins of Music*, eds S. Brown, B. Merker, and N. L. Wallin (Cambridge, MA: MIT Press), 49–63.

Slater, P. (2011). *Bird Song and Language*. Oxford: Oxford University Press, doi: 10.1093/oxfordhb/9780199541119.013.0008

Snowdon, C. T. (2003). "Expression of emotion in nonhuman animals," in *Handbook of Affective Sciences*, eds R. J. Davidson, K. R. Scherer, and H. H. Goldsmith (Oxford, UK: Oxford University Press), 457–480.

Snowdon, C. T., and Teie, D. (2013). "Emotional communication in monkeys: music to their ears?," in *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford, UK: Oxford University Press), 133–154.

Snowdon, C. T., Zimmermann, E., and Altenmüller, E. (2015). Music evolution and neuroscience. *Prog. Brain Res.* 217, 17–34. doi: 10.1016/bs.pbr.2014.11.019 doi: 10.1016/bs.pbr.2014.11.019

Sodgerel, T. (2012). "Herder's flute tsuur in traditional culture of Western Mongolia [Пастушечья флейта цуур в традиционной культуре Западной Монголии]," in *Culture of Mongol-speaking peoples in globalized space. International scientific forum, Elista, 24-27 October, 2012 [Культура монголоязычных народов в глобализирующемся пространстве]*, ed. K. Nadneyeva (Elista: Kalmyk State University), 60–61.

Sodgerel, T. (2016). Patterns of playing the Mongolian longitudinal flute: features of the repertoire. [Наигрыши на монгольской продольной флейте: особенности репертуара]. *Bull. Saint Petersburg State Instit. Cult.* 3, 126–129.

Soffer, O., Adovasio, J. M., and Hyland, D. C. (2000). The "Venus" figurines. *Curr. Anthropol.* 41, 511–537. doi: 10.1086/317381

Sørensen, L., and Karg, S. (2014). The expansion of agrarian societies towards the north - new evidence for agriculture during the Mesolithic/Neolithic transition in Southern Scandinavia. *J. Archaeol. Sci.* 51, 98–114. doi: 10.1016/j.jas.2012.08.042 doi: 10.1016/j.jas.2012.08.042

Stebbins, W. C., and Moody, D. B. (2011). "How monkeys hear the world: auditory perception in nonhuman primates," in *Comparative Hearing: Mammals. Springer Handbook of Auditory Research*, eds R. R. Fay and A. N. Popper (New York, NY: Springer), 97–133. doi: 10.1007/978-1-4612-2700-7_4

Stépanoff, C., Marchina, C., Fossier, C., and Bureau, N. (2017). Animal autonomy and intermittent coexistences: north asian modes of herding. *Curr. Anthropol.* 58, 57–81. doi: 10.1086/690120

Straehley, I. C., and Loebach, J. L. (2014). The influence of mode and musical experience on the attribution of emotions to melodic sequences. *Psychomusicology* 24, 21–34. doi: 10.1037/pmu0000032

Stricklin, W. R. (2001). "The evolution and domestication of social behavior," in *Social Behavior in Farm Animals*, eds J. L. Keeling and H. W. Gonyou (Wallingford: CABI publishing), 83–110. doi: 10.1079/9780851993973.0083

Stuart, C. K. (2008). *Life and Marriage in Skya Rgya, a Tibetan Village - Blo Brtan Rdo Rje*. New York, NY: YBK Publishers.

Suisman, D. (2009). *Selling Sounds: The Commercial Revolution in American Music*. Cambridge, MA: Harvard University Press.

Sundberg, J. (1987). *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press.

Sutton, R. A. (1991). *Traditions of Gamelan Music in Java: Musical Pluralism and Regional Identity*. Cambridge, MA: Cambridge University Press.

Svensson, E. (2015). "Upland living. The scandinavian shielings and their European Sisters," in *Nordic Middle Ages-Artefacts, Landscapes and Society. Essays in Honour of Ingvild Øye on her 70th Birthday*, eds I. Baug, J. Larsen, and S. S. Mygland (Berg: University of Bergen), 289–300.

Swain, J. P. (2002). *Harmonic Rhythm: Analysis and Interpretation*. Oxford, UK: Oxford University Press.

Swanwick, K., Tillman, J., and Maccoby, E. E. (1986). The sequence of musical development: a study of children's composition. *Br. J. Music Educ.* 3, 305–339. doi: 10.1017/S0265051700000814

Tagg, P. (2012). *Music's Meaning: A Modern Musicology for Non-Musos*. Larchmont, NY: Mass Media's Scholar's Press.

Tallerman, M. (2013). Join the dots: a musical interlude in the evolution of language? *J. Linguist.* 49, 455–487. doi: 10.1017/S0022226713000017

Tarasti, E. (1994). *A Theory of Musical Semiotics*. Indianapolis, IN: Indiana University Press.

Tarasti, E. (2012). *Semiotics of Classical Music: How Mozart, Brahms and Wagner Talk to Us*. Berlin: Mouton de Gruyter.

Tarasti, E. (ed.) (1995). *Musical Signification: Essays in the Semiotic Theory and Analysis of Music*. New York, NY: Mouton de Gruyter.

Tarr, B., Launay, J., and Dunbar, R. I. M. (2014). Music and social bonding: "Self-other" merging and neurohormonal mechanisms. *Front. Psychol.* 5:1096. doi: 10.3389/fpsyg.2014.01096

Tchotchkina, M. P. (2003). *Altaic Children Folklore [Алтайский детский фольклор]*. Gorno-Altaisk: Gorno-Altaisk state university press.

Tesch, S. (1992). The long-term development of a settlement region on the coastal plain - the Köpinge area. *Acta Archaeol. Lundensia* 4, 283–344.

Thiemel, M. (2001). Dynamics. *New Grove Diction. Music Music.* doi: 10.1093/gmo/9781561592630.article.08458

Thompson, N. S., Abbey, E., Wapner, J., Logan, C., Merritt, P. G., and Pooth, A. (2000). Variation in the bout structure of Northern Mockingbird (*Mimus polyglottos*) singing. *Bird Behav.* 13, 93–98.

Tilley, C. (1996). *An ethnography of the Neolithic: Early Prehistoric Societies in Southern Scandinavia*. Cambridge, MA: Cambridge University Press.

Titon, J. T. (2015). Ethnomusicology as the study of people making music. *Muzikoloski Zbornik* 51, 175–185. doi: 10.4312/mz.51.2.175-185

Titze, I. R. (1988). A framework for the study of vocal registers. *J. Voice* 2, 183–194. doi: 10.1016/S0892-1997(88)80075-4

Titze, I. R., Blake, D., and Wodzak, J. (2018). Intelligibility of long-distance emergency calling. *J. Voice* 32, 1–9. doi: 10.1016/J.JVOICE.2018.08.008

Tiukhteneva, S. P. (2017). On pastoral magic amongst contemporary Altaians [О скотоводческой магии у современных алтайцев]. *Oriental Stud.* 34, 62–70.

Todd, N. P. M. (1992). The dynamics of dynamics: a model of musical expression. *J. Acoust. Soc. Am.* 91, 3540–3550. doi: 10.1121/1.402843

Touma, H. H. (1996). *The Music of the Arabs, Portland*. Oregon: Amadeus Press.

Trainor, L. J. (2010). The emotional origins of music. *Phys. Life Rev.* 7, 44–45. doi: 10.1016/j.plrev.2010.01.010

Trainor, L. J., and Hannon, E. E. (2013). "Musical development," in *Psychology of Music*, ed. D. Deutsch (New York: Academic Press), 423–498.

Trehub, S. E. (2008). Music as a dishonest signal. *Behav. Brain Sci.* 31, 598–599. doi: 10.1017/S0140525X08005530

Trehub, S. E., and Hannon, E. E. (2006). Infant music perception: domain-general or domain-specific mechanisms? *Cognition* 100, 73–99. doi: 10.1016/j.cognition.2005.11.006

Trehub, S. E., Unyk, A. M., and Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant Behav. Dev.* 16, 285–295. doi: 10.1016/0163-6383(93)80036-8

Tresset, A., and Vigne, J.-D. (2011). Last hunter-gatherers and first farmers of Europe. *C. R. Biol.* 334, 182–189. doi: 10.1016/J.CRVI.2010.12.010 doi: 10.1016/j.crvi.2010.12.010

Trevarthen, C. (2019). "The psychobiology of the human spirit," in *Early Interaction and Developmental Psychopathology. Infancy*, eds G. Apter, E. Devouche, and M. Gratier (Cham: Springer International Publishing), 1–18. doi: 10.1007/978-3-030-04769-6_1

Trost, W. J., Labbé, C., and Grandjean, D. (2017). Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia* 96, 96–110. doi: 10.1016/J.NEUROPSYCHOLOGIA.2017.01.004

Tull, H. W. (1996). The killing that is not killing: men, cattle, and the origins of non-violence (ahiṃsā) in the Vedic sacrifice. *Indo Iranian J.* 39, 223–244. doi: 10.1163/000000096790085150

Tuniz, C., Bernardini, F., Turk, I., Dimkaroski, L., Mancini, L., and Dreossi, D. (2012). Did Neanderthals play music? X-ray computed micro-tomography of the divje babe 'Flute.'. *Archaeometry* 54, 581–590. doi: 10.1111/j.1475-4754.2011.00630.x

Turino, T. (2014). Peircean thought as core theory for a phenomenological ethnomusicology. *Ethnomusicology* 58, 185–221. doi: 10.5406/ethnomusicology.58.2.0185

Turk, I. (2014). *DIVJE BABE I. Upper Pleistocene Palaeolithic site in Slovenia. Opera Inst*. Ljubljana: Collegium Graphicum.

Turk, I., Dirjec, J., Bastiani, G., Pflaum, M., Lauko, T., Cimerman, F., et al. (2001). New analyses of the "flute" from Divje babe I (Slovenia) [Nove analize "piscali" iz Divjih bab I (Slovenija)]. *Arheoloski Vestnik* 52, 25–79.

Uspensky, B. (1995). *Semiotics of Art [Семиотика искусства]*. Moscow: Languages of Russian Culture.

Uttman, M. T. (2002). Eine untersuchung der teiltonspektren bei kulning und lockruftechniken anhand von beispielen aus Schweden und Finnland. *Swedish J. Music Res.* 5, 1–18.

Val'kova, V. B. (1992). *Musical Thematicism, Cognition, Culture [музыкальный тематизм, мышление, культура]*. Nizhnii Novgorod: State Nizhegorodskii University.

Vashkevich, N. (2006). *Semantics of Musical Speech. Musical Syntax [Семантика музыкальной речи. Музыкальный синтаксис]*. Tver: Musorgsky Musical College.

Ventsel, A. (2006). Hunter–herder continuum in anabarski District, NW Sakha, Siberia, Russian Federation. *Nomadic Peoples* 10, 68–86. doi: 10.3167/np.2006.100205

Vorren, Ø (1973). "Some trends of the transition from hunting to nomadic economy in Finnmark," in *Circumpolar Problems: Habitat, Economy and Social Relations in the Arctic*, ed. G. Berg (Oxford, UK: Pergamon Press), 185–194. doi: 10.1016/b978-0-08-017038-1.50026-1

Walker, R. (1997). Visual metaphors as music notation for sung vowel spectra in different cultures. *J. New Music Res.* 26, 315–345. doi: 10.1080/09298219708570733

Wallin, N. L. (1983). Pitch perception as expression for exogene and endogene coordinated oscillations. *World Music* 25, 46–64.

Wallin, N. L. (1991). *Biomusicology: Neurophysiological, Neuropsychological, and Evolutionary Perspectives on the Origins and Purposes of Music*. Hillsdale, NY: Pendragon Press.

Waser, P. M., and Waser, M. S. (1977). Experimental studies of primate vocalization: specializations for long-distance propagation. *Zeitschrift Tierpsychol.* 43, 239–263. doi: 10.1111/j.1439-0310.1977.tb00073.x doi: 10.1111/j.1439-0310.1977.tb00073.x

Welch, G. F. (2006). "Singing and vocal development," in *The Child as Musician: A Handbook of Musical Development*, ed. G. McPherson (New York, NY: Oxford University Press), 311–330. doi: 10.1093/acprof:oso/9780198530329.003.0016 doi: 10.1093/acprof:oso/9780198530329.003.0016

Wermke, K., Leising, D., and Stellzig-Eisenhauer, A. (2007). Relation of melody complexity in infants' cries to language outcome in the second year of life: a longitudinal study. *Clin. Linguist. Phonet.* 21, 961–973. doi: 10.1080/02699200701659243

Wermke, K., and Mende, W. (2009). Musical elements in human infants' cries: in the beginning is the melody. *Music. Sci.* 13, 151–175. doi: 10.1177/102986490901300208l

Whiten, A. (2011). The scope of culture in chimpanzees, humans and ancestral apes. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 997–1007. doi: 10.1098/rstb.2010.0334

Whittaker, J. C. (1994). *Flintknapping: Making and Understanding Stone Tools. Austin*. Texas: University of Texas Press.

Whittle, M. (2007). *Gait Analysis: An Introduction*. Oxford UK: Butterworth-Heinemann.

Wildgen, W. (2004). *The Evolution of Human Language: Scenarios, Principles, and Cultural Dynamics*. Philadelphia: John Benjamins Publishing.

Wiley, R. H. (1983). "The evolution of communication: information and manipulation," in *Animal Behavior: Communication*, eds T. R. Halliday and P. J. B. Slater (London: Blackwell Publishers), 156–189.

Winnington-Ingram, R. P. (2015). *Mode in Ancient Greek Music*. Cambridge, MA: Cambridge University Press.

Wulstan, D. (1971). The origin of the modes. *Stud. Eastern Chant* 2, 4–20.

Wyatt, S. (2016). "Musiqualia and vultural adaptation," in *Studien zur Musikarchäologie. Klang – Objekt – Kultur – Geschichte*, eds R. Eichmann, L.-C. Koch, and J. Fang (Berlin: Verlag Marie Leidorf GmbH), 169–194.

Yemelyanov, V. (2000). *Voice Development: Coordination and Training [Развитие эолоса: Координация и тренаж]*. Saint Petersburg: Lan.

Yip, M. J. (2006). The search for phonology in other species. *Trends Cogn. Sci.* 10, 442–446. doi: 10.1016/J.TICS.2006.08.001

Yoon, S. (2018). What's in the Song? Urtyn duu as sonic "Ritual" among mongolian herder-singers. *MUSI Cultures* 45, 92–111.

Yoshida, S., and Okanoya, K. (2005). Evolution of turn-taking: a bio-cognitive perspective. *Cogn. Stud.* 12, 153–165. doi: 10.11225/jcss.12.153

Yust, J. (2018). *Organized Time: Rhythm, Tonality, and Form*. Oxford, New York: Oxford University Press.

Zabolotskaya, P. Y. (2009). *Folkloric Theater of Yakuts: An Attempt of Historic-Hystrionic Research [Фольклорный театрякутов (опыт историко-театроведческоэо исследования)]*. Ulan-Ude: East Siberian state academy of culture and arts.

Zeder, M. A. (2008). Domestication and early agriculture in the mediterranean basin: origins, diffusion, and impact. *Proc. Natl. Acad. Sci. U.S.A.* 105, 11597–11604. doi: 10.1073/pnas.0801317105

Zemtsovsky, I. (1979). "On creative nature of folklore [О творческой природе фольклора]," in *Stylistic Trends in Soviet Music 1960-70s [ Сти левые т енденции в Совет ской музыке 1970-70 эодов]*, ed. A. N. Kriukov (Leningrad: Lenuprizdat), 137–147.

Zemtsovsky, I. (1983). "Song as a historic phenomenon [Песня как исторический феномен]," in *Popular Song: Problems of Study [Народная песня. Проблеумы изучения]*, ed. V. Gusev (Leningrad: State Academy of Theatric Arts), 22–35.

Zemtsovsky, I. (1997). An attempt at a synthetic paradigm. *Ethnomusicology* 41, 185–205. doi: 10.2307/852602

Zemtsovsky, I. (2002). The apology of text [Апология текста]. *Music. Acad.* 4, 100–110.

Zemtsovsky, I. (2005). Neither East nor West; in between but not a bridge: A riddle for a new discipline, the ethnogeomusicology. *Muzikologija: Časopis Muzikološkog Instituta Srpske Akademije Nauka i Umetnosti* 5, 195–203. doi: 10.2298/MUZ0505195Z

Zemtsovsky, I., and Kunanbayeva, A. (2011). "Homo Lyricus, or lyrical song in ethnomusicological stratigraphy of 'folk culture' [Homo Lyricus, или лирическая песня в ƨтномузыковедческой стратитрафии «фольклорной культуры»]," in *Classic folklore today. Proceedings of the conference "90 years anniversary of B.N.Putilov's birthday*, (Sankt-Petersburg: Dmitri Bulanin), 199–249.

Zeskind, P. S. (1985). "A developmental perspective of infant crying," in *Infant Crying: Theoretical and research perspectives*, eds C. F. Z. Boukydis and B. M. Lester (Boston, MA: Springer), 159–185. doi: 10.1007/978-1-4613-2381-5_8 doi: 10.1007/978-1-4613-2381-5_8

Zilhão, J. (2014). "The upper palaeolithic of Europe," in *The Cambridge World Prehistory*, eds C. Renfrew and P. G. Bahn (Cambridge, MA: Cambridge University Press), 1753–1785.

Zimmermann, E., Leliveld, L., and Schehka, S. (2013). "Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices," in *Evolution of Emotional Communication*, eds E. Altenmüller, S. Schmidt, and E. Zimmermann (Oxford, UK: Oxford University Press), 116–132. doi: 10.1093/acprof:oso/9780199583560.003.0008

Zuberbühler, K. (2017). "The primate roots of human language," in *Primate Hearing and Communication*, eds R. M. Quam, M. Rosa, and J. L. Arsuaga (Berlin: Springer), 175–200. doi: 10.1007/978-3-319-59478-1_7

Zubrow, E. B. W., and Blake, E. C. (2006). "The origin of music and rhythm," in *Archaeoacoustics*, eds C. Scarre and G. Lawson (Cambridge, MA: McDonald Institute for Archaeological Research), 142.

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2020 Nikolsky. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*

# Advantages of publishing in Frontiers

**OPEN ACCESS**
Articles are free to read for greatest visibility and readership

**FAST PUBLICATION**
Around 90 days from submission to decision

**HIGH QUALITY PEER-REVIEW**
Rigorous, collaborative, and constructive peer-review

**TRANSPARENT PEER-REVIEW**
Editors and reviewers acknowledged by name on published articles

**Frontiers**
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

**Visit us:** www.frontiersin.org
**Contact us:** info@frontiersin.org | +41 21 510 17 00

**REPRODUCIBILITY OF RESEARCH**
Support open data and methods to enhance research reproducibility

**DIGITAL PUBLISHING**
Articles designed for optimal readership across devices

**FOLLOW US**
@frontiersin

**IMPACT METRICS**
Advanced article metrics track visibility across digital media

**EXTENSIVE PROMOTION**
Marketing and promotion of impactful research

**LOOP RESEARCH NETWORK**
Our network increases your article's readership