Inverse and indirect mappings in embodied AI systems in everyday environments

Maham Riaz[1]*, Çağrı Erdem[2] and Alexander Refsum Jensenius[1]

[1]Department of Musicology, RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Oslo, Norway, [2]Department of Informatics, University of Oslo, Oslo, Norway

This paper explores how musicking technologies—interactive systems with musical properties—can enhance everyday public environments. We are particularly interested in investigating the effects of musical interactions in non-musical settings, such as offices, meeting rooms, and social work areas. Traditional music technologies (such as instruments) are built for goal-directed, conscious, and voluntary interactions. We propose a new perspective on embodied AI through systems that utilize indirect, inverse, unconscious, and, at times, involuntary interactions. Four different sound/music systems are examined and discussed with regard to their activity level: a reactive "birdbox," a reactive painting, active self-playing guitars, and interactive music balls. All these systems are multimodal, containing sensors that detect various physical inputs to produce sound and light, and having varying levels of perceived agency. The paper explores differences between direct/indirect and regular/inverse embodied AI paradigms. This study demonstrates how minimalistic interactions have the potential to yield complex and engaging musicking experiences, challenging the norms of overly intricate AI implementations.

1 Introduction

Recent technological developments have opened new possibilities for integrating sound and music into everyday environments in unconventional ways. Musical instruments are typically designed for conscious, deliberate interaction, where users intentionally manipulate devices to produce (musical) sound (Jensenius, 2022). Many of our everyday physical and digital interactions are structured around functional engagement rather than musical expression. They exist as direct exchanges within a structured environment, forming predictable interaction cycles. Everyday interactions are typically direct and straightforward engagements with objects in a space, characterized by a low level of complexity, thereby resulting in low cognitive loads. For example, opening a door requires only mechanical force applied to a handle; the system does not infer intent or modify its response based on past interactions.

As part of our ongoing investigations into subtle aesthetic experiences in everyday environments (Riaz et al., 2025a,b), we are curious about how new technologies can allow musical interactions that are indirect, unconscious, and even involuntary and not intended to be musical in the first place. Such systems can transform everyday spaces, such as offices, meeting rooms, and other public or semi-public areas, into environments that playfully respond to human activity.

Research on new interfaces for musical expression (Wessel and Wright, 2002; Jensenius and Lyons, 2017), sonic interaction design (Franinovic and Serafin, 2013; Geronazzo and Serafin, 2023; Jensenius, 2024), and embodied interfaces (Lesaffre et al., 2017; Holland et al., 2019) has demonstrated that technology can expand musical expression and support new forms of participatory experiences. For instance, systems that convert bodily actions and gestures, environmental changes, or incidental motion into sound have been proposed to broaden creative expression beyond traditional performance contexts (Tanaka, 2006; Ichino and Nao, 2018; Headlee et al., 2010). These approaches highlight a growing interest in interactions that are not strictly goal-directed or intentional but emerge naturally from engagement with the environment, often originating from non-musical actions and contexts.

In this paper, we propose an alternative interaction paradigm to the design of embodied AI systems, focusing on indirect and inverse mappings. This can be thought of as a type of *musicking technology*, essentially an interactive system designed to incorporate musical properties without the direct interaction typical in traditional music technologies. This approach aligns well with trends in ubiquitous computing and ambient interaction design, which aim to make technology accessible from the periphery of our attention rather than its center (Gaye et al., 2006; Tanaka, 2006). By responding to non-musical actions, such as routine activities, locomotion, or changes in environmental conditions, these systems encourage spontaneous participation, often without even realizing it.

Our approach can be thought of as "techno-cognitive" (Jensenius, 2022), integrating human perception and cognition theories into technological development. Developing musical instruments and other interactive music systems provides an excellent testbed for cognitive theories since they allow non-harmful and playful experimentation. We can also learn a great deal about human cognition by prototyping new instruments that, in various ways, engage our senses. Thus, prototyping and reflecting on interactive systems is a good example of how the "logies" of musico-logy, techno-logy, and psycho-logy play together.

This paper examines how musicking technologies with indirect and inverse interaction properties can foster engagement and creativity in everyday, non-musical settings. We begin by discussing some key concepts before laying out an embodied AI interaction framework. This is followed by presenting four devices that in various ways explore parts of this framework: (1) a reactive "birdbox" that plays back a soundfile when people pass before the device, (2) a reactive painting that responds with sound and light to passers-by, (3) a set of self-playing guitars that listen to each other and inversely respond to humans, (4) interactive balls that sense their environment and encourage touch and motion (Figure 1). These devices all challenge the assumption that meaningful interactions in AI-based systems require complex intelligence with clearly defined goals. We aim to demonstrate that minimalist systems can generate engaging musicking experiences from incidental, unconscious, or entirely non-musical interactions.

2 Key concepts

This section lays the groundwork by defining several key concepts essential to understanding our framework for embodied AI systems in everyday environments. Our approach is founded on theories of embodied music cognition (Leman, 2007), positing that human cognition is based on an action–perception loop between mind and body, as well as multisensory interactions between agents, objects, and the environment (Figure 2).
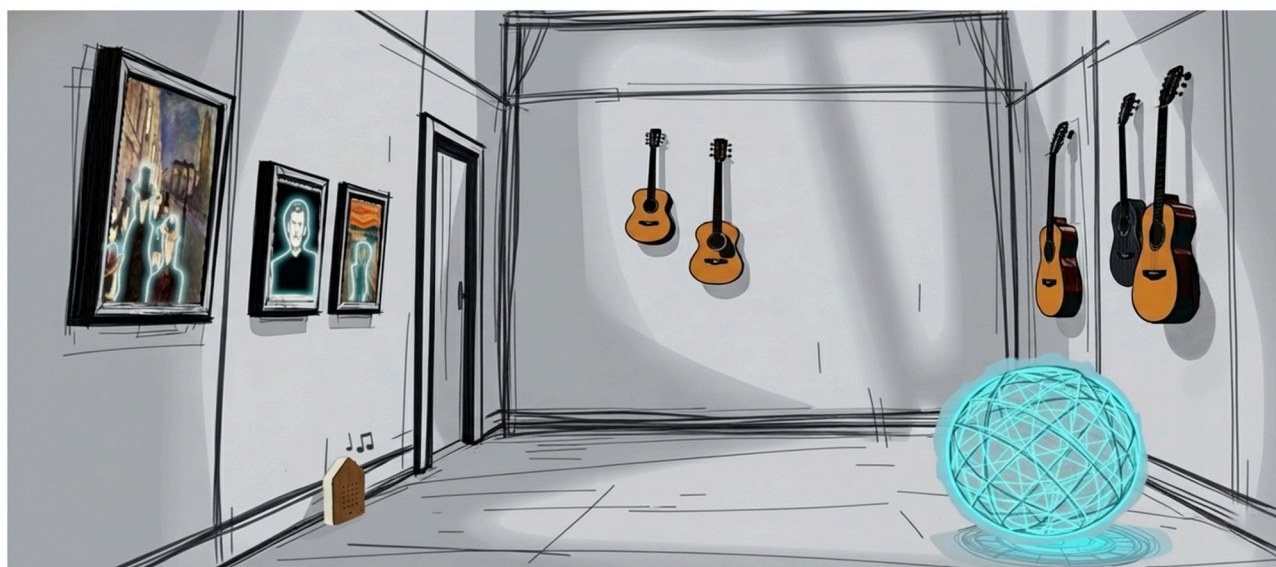

FIGURE 1
The four use cases considered in this article (from left): reactive paintings, a birdbox, self-playing guitars, and an interactive music ball.

2.1 Embodied cognition

Embodied cognition offers an understanding of musical experience as a process involving bodily action, perceptual feedback, social context, and environmental influences. Specifically, embodied music cognition can be thought of as a layered process involving (a) action–perception loops between body and mind; (b) actions, reactions, and interactions between performers and perceivers; (c) action–reaction loops between performers and their instruments; and (d) interactions among performers, perceivers, and the environment (Jensenius, 2022).

Action–perception loops describe the continuous, reciprocal relationship in which motor systems influence and are influenced by sensory systems (Gibson, 1979). The common-coding theory states that perception and action share representational codes in the brain (Maes et al., 2014). Observing or hearing a movement can activate corresponding motor plans, and motor planning can shape perception through forward models (which predict the sensory outcomes of actions) or inverse models (which infer the actions that produced a given sensory input). Empirical research in music cognition shows that listeners' motor systems engage during the perception of rhythmic or expressive features, and that altering a listener's posture or movement can affect their perception of rhythmic stability, tempo, or expressive timing (Maes et al., 2014).

Performers experience what can be thought of as action–reaction loops with their instruments. For example, for a violinist, the action of moving the bow produces sound through string vibration; this auditory output, combined with tactile feedback from the instrument and proprioceptive feedback from the player's joints and muscles, informs subsequent motor adjustments. This is a recurring loop: the performer anticipates the sonic results of specific musical gestures, listens to the actual sound produced, senses tactile and vibrational feedback from the instrument, and adapts accordingly (Mice and McPherson, 2021).

Enactive approaches to embodied cognition also emphasize the social dimension in interactions between performers and perceivers, as well as the broader relationships between individuals and their environment. For example, in ensemble performance, coordination among performers through visual cues, gestures, and timing requires continuous mutual adjustments. In a study on clarinet and piano duos (Bishop et al., 2019), the availability of visual contact increased gestural synchrony and the use of visual signals, particularly during passages with temporal irregularity; when visual contact was removed, coordination suffered.

On the perceiver side, audiences in concert settings exhibit bodily responses, such as changes in breathing, posture, and subtle movements, that correlate with subjective experiences of absorption and emotional affect (Maes et al., 2014; Dell'Anna et al., 2021). Recent studies using motion capture, physiological measurements, and interviews suggest that listeners create meaning through their bodies mediated by the performance environment, including the stage, acoustics, visual cues, and the presence of others, as well as their own musical skills (Haswell-Martin et al., 2025).

These processes exist through multimodal perception and sensorimotor integration. While primarily an auditory art,
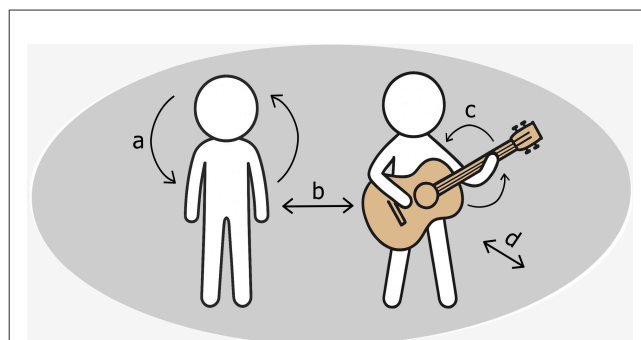

FIGURE 2
Embodied music cognition is a multilayered process: (a) action—perception loops between the body and mind, (b) actions, reactions, and interactions between performers and perceivers, (c) action—reaction loops between performers and their instruments, and (d) interactions between the performers/perceivers and the environment.

music is frequently experienced in contexts rich with visual information—performers' gestures, lighting, digital visuals, or even album art—all of which may shape interpretation and response, making it a "bimodal" endeavor. For example, a performer's expressive body language can alter listeners' judgments of musical emotion. In one study, participants' interpretations of emotionally expressive body movements were biased when accompanied by congruent or incongruent instrumental music, showing how auditory input modulates visual meaning (Van Den Stock et al., 2009).

Research on crossmodal interaction further shows how different sensory channels interact to produce musical experience. Visual information from performers' movements systematically shapes listeners' judgments of phrasing and tension, with combined auditory–visual presentations producing stronger emotional effects than either modality alone (Vines et al., 2006). Visual information about gesture duration alters the perceived length of a sound, illustrating how motor cues bias auditory evaluation (Schutz and Lipscomb, 2007). Listeners draw not only on auditory cues but also on visual, motor, and even narrative associations when anticipating musical events (Godøy et al., 2006; Caramiaux et al., 2010).

Processes related to motor imagery and mirror neurons enable listeners to perceive expressive movement even when they themselves are stationary (Maes et al., 2014). The multimodal and crossmodal channels feed into action–perception loops for both performers, who see and feel their own actions, and listeners, who observe, hear, and sometimes internally simulate actions. Therefore, a more embodied understanding of music emerges through the integration of auditory, visual, tactile, and motor domains.

2.2 Musicking technology

In his seminal book on embodied music cognition (Leman, 2007), Marc Leman used the term "mediation technology" to

describe various digital systems that support musical exploration. The *D-Jogger*, developed at IPEM (Moens, 2018), is an example of such a musical mediation technology. It is an "active listening" device that adjusts the tempo of the music to the user's running speed. A musical experience is at the core, and the user can indirectly control one musical feature (the tempo) through the designed mapping from running speed.

Our use of the term "musicking technology" combines Leman's mediation technology and Christopher Small's concept "musicking" (Small, 1998). Small argues that music is not a "thing" but something you do; hence, music should be used as a verb ("to music"). This is particularly important when considering new musical instruments and interfaces (Jensenius and Lyons, 2017), which often have very different qualities from the traditional ones. The term also inherently encompasses situations that are not based on active music-making, including interactions that are non-active or not musically intended. We highlight this aspect as it allows us to frame both deliberate musical practices and incidental sonic interactions within the same perspective.

The field of interaction design emphasizes the distinction between incidental and intentional feedback (Dourish, 2001). Intentional feedback is designed explicitly to convey information to the user. For example, when you click a touchscreen button that changes color or vibrates, the system provides feedback to confirm your action. Incidental feedback is a natural consequence of an interaction, but it is not explicitly designed to convey a specific message. In the case of a physical button, the resistance and click sound are not necessarily designed as feedback, but they provide valuable cues about the action when the button is pressed. Everyday interactions, such as opening doors, pressing elevator buttons, or typing on a keyboard, generate incidental sounds. This sound differs from the sound-producing actions found in musical instruments and other devices designed for structured musical interactions (Jensenius, 2022). The sounds generated through everyday interactions can be seen as a type of sonic feedback, providing cues about action through mechanical byproducts rather than a deliberate mapping of action to sound.

Interestingly, many digitally based products integrate speakers and sound designs that mimic the sonic feedback from mechanical devices. One example is the clicking sounds when typing on mobile phones. A design challenge, however, is that users can opt to disable such sonic feedback. Nowadays, many screen-based devices provide multiple feedback channels, including visual, sonic, and haptic. In addition to such designed feedback, there may be acoustic feedback caused by, for example, fingernails hitting a mobile phone screen. These examples show how sonic interaction extends beyond deliberate design, where purposeful and incidental sound production mix together.

This distinction resonates with the design of musicking technologies. Traditional (acoustic) instruments are built around action–sound couplings defined by the mechanical and acoustical properties of objects, which give each instrument its characteristic set of sonic affordances (Jensenius, 2022). Interactive music systems that use digital computing, however, are based on specific action–sound mappings, input actions, and a sound engine, which may be intuitive or non-intuitive, direct or indirect, conscious or unconscious, voluntary or involuntary. Many digital products occupy a space between these extremes, deliberately mimicking

the incidental sonic feedback of mechanical devices; for example, the simulated clicking sounds of mobile phone keyboards. Yet users can often disable such sounds, revealing the tension between designed (intentional) and incidental forms of sonic feedback in everyday technologies.

For a performer on stage, conscious, voluntary, and direct mapping is usually better, as it makes more sense to both performers and perceivers. On the other hand, when developing interactive systems for everyday environments, we do not aim to create something that is immediately obvious. We are interested in exploring how subtle interactions, with both intentional and incidental feedback, can improve everyday environments.

Advanced sonification systems can be considered a type of musicking technology. Such systems are often designed to convey meaningful information, such as energy expenditure or air pollution levels. Including meaningful sonification is also part of our research agenda. For example, projects like *Sound for Energy* have tested how sound can make energy consumption patterns more perceptible, promoting awareness through auditory feedback (Pauletto, 2024). However, in this context, we are more interested in understanding the subtleties, limitations, and possibilities of creating systems on the edge between musicking and music technologies, where the distinction between designed mappings and incidental sonic byproducts becomes blurred.

2.3 Musical AI

Artificial intelligence (AI) encompasses a broad range of techniques that enable machines to perform tasks that require some level of "intelligence," from pattern recognition to decision-making. Two main approaches to AI are rule-based and learning-based systems. One could say that rule-based AI is the "classic" approach, which has lost popularity recently because machine learning, especially deep learning, has dominated the media and public discourse. However, both rule-based and learning-based approaches have their pros and cons.

Rule-based AI systems use predefined logic, applying "if–then" rules to process inputs. Such rule-based AI systems offer simplicity and can be scaled to accommodate incremental complexity as desired by the maker (Russell and Norvig, 2016). Historically, composers have employed structured methods, such as combinatorial techniques and musical automata, recognizing the inherently rule-driven nature of Western music theory (Erdem, 2022; Franinovic and Serafin, 2013; Miranda, 2021). Algorithmic composition emerged in the 1950s (Sandred, 2021), exemplified by the *Illiac Suite* by Hiller and Isaacson (Hiller and Isaacson, 1993). Later, David Cope's *Experiments in Musical Intelligence (EMI)* generated music in the style of various composers using similar methods (Cope, 2000).

Rule-based systems were foundational in AI, particularly in expert systems (Barr et al., 1981). However, the rise of statistical and machine-learning approaches led to a shift in AI research priorities (Goodfellow et al., 2016). Despite this, rule-based interactive AI continues to be influential in music. Software like *Band-in-a-Box* generates accompaniment tracks using predefined music theory rules (Beauchamp-Williamson, 2004). *SmartMusic* aids

music education by providing real-time performance feedback (Owen, 2015). George Lewis' *Voyager* engages in improvisational performance with musicians through rule-based decision-making (Steinbeck, 2018).

In his framework for understanding how rule-based systems can structure interactive music-making, Robert Rowe differentiates between "performance-driven" and "composition-driven" systems (Rowe, 1992). Each is defined by how rules are applied to interpret and respond to user input. Performance-driven systems, such as gesture-responsive instruments, are particularly relevant to scenarios involving indirect or unconscious interactions—a concept this paper applies to non-musical interactions in everyday environments.

Learning-based AI systems recognize patterns, make predictions, and self-improve over time using data-driven methods (Vigliensoni and Fiebrink, 2025). Unlike rule-based AI, which follows pre-defined instructions, these systems adjust internal parameters to optimize performance (Goodfellow et al., 2016). Neural networks exemplify this, learning patterns without explicit programming (LeCun et al., 2015). Machine learning, deep learning, and reinforcement learning are subcategories of learning-based AI (Russell and Norvig, 2016). The narrative around AI has diverted toward more complex data-driven methods, leading people to associate it primarily with machine learning (Mitchell, 2019). This approach has advanced fields like natural language processing (Manning et al., 2008), image recognition (Krizhevsky et al., 2017), and autonomous systems (Sutton and Barto, 2018).

However, AI is not limited to data-driven learning. Many practical systems still rely on rule-based or hybrid models (Bouchard et al., 2016; Shortliffe, 2012; Groover, 2016; Russell and Norvig, 2016). Interesting musical interactions can arise from combining rule-based and learning-based systems in musical AI.

2.4 Active, reactive, and interactive systems

When considering various types of (embodied) AI systems, we see the need to differentiate between "active," "reactive," and "interactive" systems. Active systems perform actions based on their internal goals or states, without requiring external stimuli. For example, an AI-based music composition system that generates music autonomously using either rule-based or generative algorithms.

Reactive Systems, on the other hand, respond to external stimuli or environmental changes. They do not have internal goals but react directly to an input without any feedback loop. A volume button or a simple key-based instrument that triggers a sample can be considered a reactive system. We would even argue that many musical instruments are reactive—and not interactive—because of the lack of a feedback loop in the system. Even though a (human) user can achieve what appears to be interaction with such a device, we suggest that this is an action–reaction chain rather than true interaction.

We reserve the term interactive system for a continuous two-way interaction in which both the user and the device respond to external stimuli and initiate actions based on the inputs. It is possible to achieve complex interactions with rule-based systems. However, this is where learning-based systems may be advantageous since they can continuously adapt their responses to input streams. However, we would refrain from saying that learning-based systems are inherently interactive. Many of today's prompt-based generative systems are reactive in the sense that they act and react without any continuous feedback loop.

2.5 Embodied musical interaction

This paper primarily explores non-tangible interactive systems that blend physical objects with both rule- and learning-based AI. Unlike traditional, everyday, direct interactions, object-based interactions enable users to engage with "intelligent," responsive objects that produce digital output based on human input.

Theories of embodied cognition suggest that human understanding and thought are shaped by bodily experiences rather than being purely abstract or mental (Wilson, 2002). When applied to object-based systems, this perspective reveals that interaction is not merely about executing commands but also about exploring, sensing, and responding through different bodily senses. This approach challenges the conventional input–output model of human–computer interaction by making the body an integral part of the process (Dourish, 2001).

A distinction must be made between embodied systems and merely embedded systems. Embedded systems focus on integrating digital functionality within physical objects (Ganssle, 2008; Henzinger and Sifakis, 2007), with limited attention to user interaction beyond simple triggers and controls. On the other hand, embodied systems engage users in interactions where the system's responses are influenced by the user's bodily presence, movement, and other sensory inputs (Abrahamson and Lindgren, 2014). Multimodal sensing and actuating are essential. Such systems require multiple input types (such as motion, light, or sound) and output (such as sound, visual cues, or haptic feedback).

3 An embodied AI interaction framework

Analyzing user influence and experience in interactive systems requires examining the specific modes of control afforded by their design. The degree to which an action directly or indirectly controls a system's output has implications for perception, agency, and engagement. The following subsections introduce the concepts of (in)direct control and inverse mappings before situating these concepts within our broader embodied AI framework.

3.1 Direct vs. indirect control

Direct control occurs when user actions immediately and explicitly affect a system's output (Shneiderman and Plaisant, 2010). This is evident when a musician presses a piano key or strums a guitar string to generate sounds. In the context of computing-based (big or small) interactive systems, direct control might involve using buttons, sliders, or touchscreens. Direct control is often characterized by the user's intentionality and the immediate

feedback provided by the system, which creates a sense of agency and mastery over the device (Norman, 2013).

Conversely, indirect control involves systems where user actions influence outcomes in less apparent ways. For example, in generative music systems, sound may be shaped by environmental sensors (such as light or motion) or algorithms, rather than through direct physical engagement and energy transfer. This type of control operates on passive, incidental engagement with no explicit user intention (Gaye et al., 2006). In indirect interaction, users do not control the system directly; instead, their actions influence an intermediary object, which in turn translates those actions into responses. For example, a motion sensor can trigger a drum sound instead of hitting a drum with a drumstick. Somewhat unconventional is inverse interaction, where reducing user input leads to increased system output (Gonzalez Sanchez et al., 2018; Manning et al., 2008). This disrupts intuition and encourages users to rethink their bodily engagement with the interactive system. An example would be a system that amplifies sound as the user reduces movement.

Some interactive systems combine direct and indirect control, allowing users to actively manipulate some elements while the system also responds to external factors. Sound installations, for example, may permit hands-on interaction and adapt to environmental conditions.

3.2 Inverse mappings

Inverse mapping in interactive objects refers to a design strategy in which the connection between the user input and device output is intentionally reversed, creating unexpected, non-linear interactions (Gonzalez Sanchez et al., 2018; Martin et al., 2018). Unlike direct input–output mappings, inverse mappings introduce a layer of abstraction that can be more engaging from the user's perspective.

Some games use inverse mapping by reversing left–right or up–down movement controls. Inverting the y-axis is especially common in flight simulators or aviation-related games (Frischmann et al., 2015). Generative music tools may invert expected sound responses, such as a synthesizer where pressing lower keys produces high-pitched sounds or when striking harder causes a softer tone, subverting conventional direct mapping schemes. Inverse interaction prioritizes exploration over immediate control. Whether using an intricate rule-based AI or a learning-based AI that adapts to the user's bodily actions over time, such indirect or inverse mappings blur the line between performer and device. They transform interaction into an iterative, evolving dialogue between humans and machines.

3.3 The framework

Our proposed embodied AI interaction framework builds on the "musicking quadrant" (Jensenius, 2022), describing four musical roles: the maker, the performer, the perceiver, and the analyst. Traditionally, there has been a clear separation between performers and perceivers. The former plays a musical instrument with direct actions. The latter sees/hears the performance without
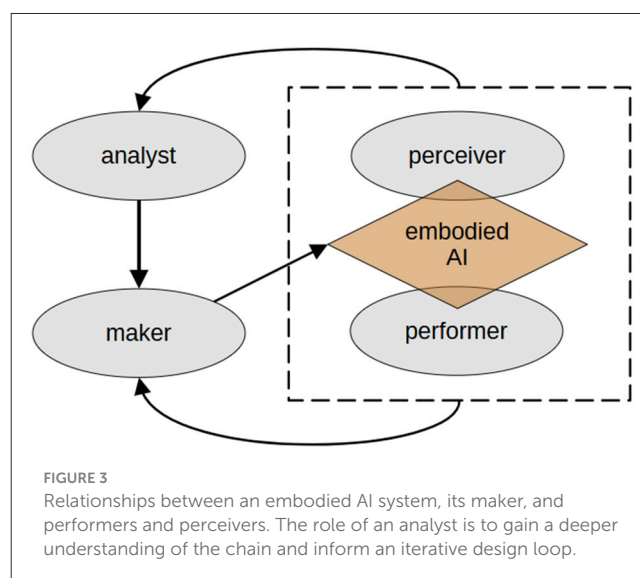
controlling the sound. New musicking technologies allow for creating devices that blur the distinction between performers and perceivers. An "active listening" device, like the D-Jogger mentioned above, is an example of a device in which the perceiver can "play" and control music indirectly through running (Moens, 2018). Various musical art installations also explore how perceivers can become performers through their motion in a space (Ahmed, 2018; Jacucci et al., 2010).

Figure 3 illustrates how an embodied AI system can be considered a bridge between performers and perceivers. Such a system can be programmed with a set of intentionalities defined by the maker, whose role is to design and create the system's interactions. The designed mappings can vary, ranging from direct to indirect, inverse, or non-inverse forms of interaction.

How a device responds to the performer's actions—whether voluntary, involuntary, conscious, or unconscious—hinges on the maker's initial intentions and the affordances embedded within the device's design. We use the term affordance to refer to a system's perceived action possibilities here. This aligns with Donald Norman's approach (Norman, 2013), which, again, was inspired by Gibson's ecological psychology (Gibson, 1977). A system's affordances define its "action potential."

The performer's engagement with a device, whether involuntary (out of reflex) or voluntary, is mediated by these affordances, which can be perceptible or imperceptible and are influenced by environmental factors such as noise levels. Affordances are inherently tied to the performer's perception and interaction with the environment, and these interactions can be further shaped by the performer's level of attention and consciousness (Gibson, 1979). Attention, in turn, makes the performer notice changes induced by their actions. If the performer is conscious of their surroundings, they may be more aware of how their motion or action triggers responses from the device. Conversely, a state of inattention may result in interactions that are less consciously perceived or even unnoticed.

The performer's level of consciousness, which is their cognitive state of mind, directly informs the degree to which they are

Relationships between an embodied AI system, its maker, and performers and perceivers. The role of an analyst is to gain a deeper understanding of the chain and inform an iterative design loop.

aware of and engaged with the device's outputs. In this regard, the performer's expectation of the device's affordances—shaped by prior experience or context—guides their interaction with the device. This reflects the concept of "perceived affordances," in which users form expectations about a device's functions based on its design (Norman, 2013). However, these expectations can evolve or be disrupted based on the actual interaction. For instance, a user might initially expect a device to respond in a certain way but may perceive a level of agency or responsiveness that diverges from their expectations, leading to shifts in their understanding of the device's affordances.

The perceiver, who observes an interaction without directly engaging with the device, plays a secondary yet significant role. While the perceiver does not experience the interaction firsthand, they are positioned to interpret and understand it from an external perspective. The perceiver's attention and consciousness, similar to those of the performer, influence how they perceive the dynamic between the performer and the device. The perceiver may notice subtleties in the interaction that the performer is unaware of, either due to the performer's inattention or lack of awareness of the device's outputs.

The analyst, positioned as an external observer, studies and interprets the interaction between the embodied AI, performer, and perceiver. The analyst's detached and critical perspective examines the cognitive and environmental factors that shape these interactions. This analysis helps inform the designer about potential refinements to the device.

4 Case studies

In the following, we examine four devices we have either developed or used in different contexts. They have been selected to exemplify several points outlined in the theoretical discussion above. This varied selection allows us to examine how different mapping strategies influence human cognitive and behavioral responses in everyday environments.

4.1 Birdbox

The birdbox is a device by Relaxound GmbH[1] that triggers bird sounds when someone passes in front of it (Figure 4). The sound source is a naturally colored oak box with a small speaker inside. The motion sensor has a range of 1.90 meters, and a 20-second bird sound sample is played as a one-shot sound, meaning it does not loop and only repeats if another motion (such as a light change) is detected. It can be seen as an example of a very simple, reactive system.

Based on the above reasoning, the birdbox has no interaction; it only reacts to a user passing by. From a user's perspective, this can be seen as an example of indirect control, as one typically triggers the sound without being aware of the device's presence. Thus, the "interaction" is unintentional, as people do not actively trigger the sounds. The bird sounds are played due to environmental changes,

———————

FIGURE 4
The audio device by Relaxound GmbH is a small box that emits bird sounds upon motion detection.

making the experience incidental and involuntary. This enables the device to coexist in everyday spaces without requiring the user to focus attention or take deliberate action.

In previous research (Riaz et al., 2025a), we report on a study of people's perception of the birdbox when placed in various public spaces within an office building. Observations and survey results from an intervention study indicated that participants primarily engaged with the bird sounds through "causal" listening (Schaeffer, 1967; Chion, 1994), attempting to identify their source, thus triggering cognitive processes. Overall, participants reacted positively to the bird sounds, appreciating the calmness and surprise they brought to the environment. Our analysis showed that the relative loudness of the sounds significantly influenced the experience; sounds that were too loud felt unpleasant, while those that were too quiet went unnoticed due to background noise. These findings underscore the importance of automatic level adjustments and considering acoustic conditions in soundscape interventions.

4.2 Reactive painting

Our prototype system *Evening on Karl Johan Street Reimagined* is based on a replica of the Norwegian painter Edvard Munch's famous work *Aften på Karl Johan* (1892), which depicts a scene on the main street in Oslo, Norway (Figure 5). The painting features a haunting, zombie-like crowd approaching the viewer, along with distinct visual elements like the silhouette of Munch himself, houses, and trees lining the street.
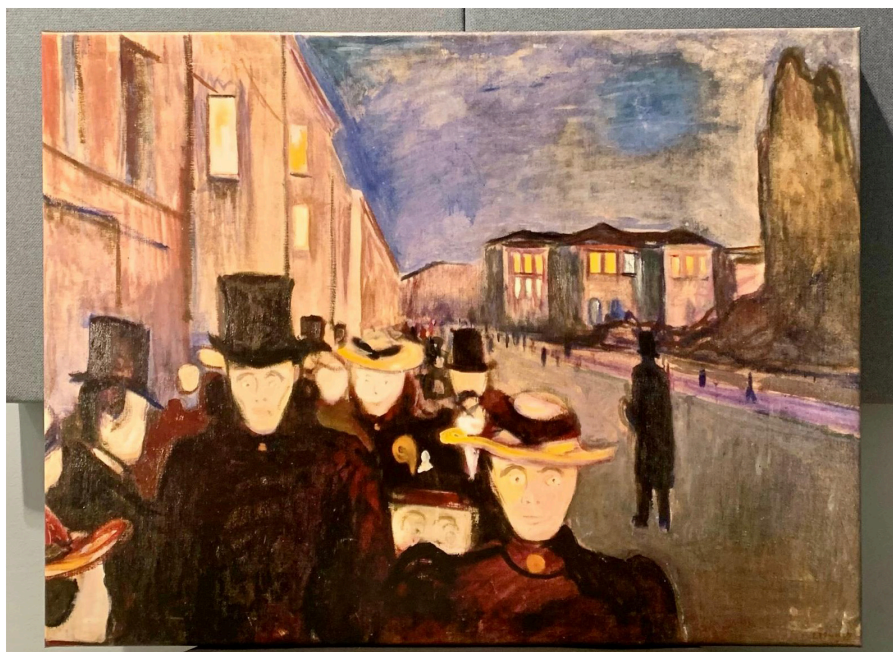
FIGURE 5
Evening on Karl Johan Street Reimagined is a reactive painting that responds with light and sound based on environmental changes and people's behavior (reproduced by the authors from Edvard Munch's "Evening on Karl Johan Street" (1892), https://commons.wikimedia.org/wiki/File: Evening_on_Karl_Johan_Street.jpg, licensed under CC0).

This project transforms the painting from a static art piece into an auditory–visual reactive installation.[2] The idea was to explore how light and sound features—subtle, multimodal responses—could reshape viewers' perceptions of the scene and foster the attribution of agency to an otherwise static object. Multiple layers of active and reactive light and sound bring the painting to life, transforming it into a multisensory storytelling experience that extends the eerie, tense atmosphere of the original artwork. Complex crossmodal links between music and paintings are generally thought to be mediated mainly through emotion (Spence and Di Stefano, 2024), which influences user engagement.

The reactive painting is a rule-based system based on predefined mappings between sensor inputs and the resulting sound and light outputs. It integrates several physical sensors, including light-dependent resistors and infrared sensors, which detect environmental changes. All the electronics are built into the frame, including a Bela board (Morreale et al., 2017) running Pure Data patches, a speaker, and light-emitting diodes (LEDs). This embedded approach is vital, as it allows the painting to be hung on a wall next to other paintings. The aim was to blend it into an environment, positively surprising people passing by with subtle light and sound changes.

The painting was designed to hang in a corridor and respond to changes in light during the day. If the light falls below a certain threshold, it triggers specific LEDs to light up behind the windows in the painting. The idea was to recreate the feeling of different times of day. Infrared sensors detect the motion of people near

the painting, triggering either sound effects (like insect, bird, or seagull sounds) or toggling the playback of ambient background sounds (like wind or crowd murmurs). The mapping follows a random selection logic so that the user will experience the painting differently each time they pass by. This unpredictability invites cognitive engagement and the ongoing search for meaning when interacting with technology in a non-musical setting.

4.3 Self-playing guitars

The self-playing guitars project features six traditional acoustic guitars augmented with embedded sensors (Gonzalez Sanchez et al., 2018) (Figure 6). Each guitar has a Bela micro-computer that produces electronic sound played through an actuator on its back. It senses its environment using an infrared distance sensor, accelerometer, and microphone. All sound is generated solely through the acoustic guitar's vibration, without the use of external speakers. Each instrument is an autonomous device with multiple sensing and actuating possibilities. They can interact with each other and also engage human users through various sensory and sound-based interactions.

Our self-playing guitar platform has been used for several different installation projects and performances (Jensenius, 2022). Several of these have been based on indirect control. The system runs on sensors, including infrared distance sensors, to detect proximity and movement. This non-traditional mode of interaction allows users to trigger soundscapes through physical presence rather than direct performance. In one mode, users standing still in front of a guitar will trigger sounds designed to mimic human

---

2   A video of the reactive painting *Evening on Karl Johan Reimagined* can be accessed at https://doi.org/10.5281/zenodo.17179396.

FIGURE 6
Six self-playing guitars hanging in a space where they can freely rotate and sense the presence of other guitars and people passing by.

breathing. The guitars challenge conventional performance ideas by limiting physical interaction to proximity and stillness, focusing instead on "microinteraction," where minimal or no motion creates sound. This is a type of inverse interaction, where actions such as motion inhibition become the primary mode of interaction, turning a standard instrument into an unconventional musical interface.

The system operates as a distributed network of independent, self-contained guitars, each with its own embedded computing platform. Over the years, we have deployed different patches on these guitars: some with rule-based, some with learning-based AI,[3] and others combining both approaches (Erdem et al., 2022). In rule-based patches, each guitar follows pre-programmed algorithms that govern how it responds to sensor input. When a user approaches within a certain distance from the guitar, the infrared sensor triggers sounds based on proximity thresholds. Deterministic rules define the sound responses, such as "if a user is still for a set period, increase the volume of the breathing sound." This combination of distributed architecture with rule-based interaction creates an autonomous system that encourages users to rethink their relationship with traditional musical instruments, transforming passive objects into interactive, responsive participants with their own sonic character in a shared sound environment.

We have also developed several different learning-based patches. One approach is based on updating a model of an environment's soundscape through sound analysis and adjusting a complex drone sound to its surroundings. Another is based on the entrainment principle, continuously adapting the frequency and phase of a pulsing sound using a Firefly-inspired algorithm

(Nymoen et al., 2014). Even though this is a very simple learning-based algorithm, complexity arises due to the presence of multiple physical objects that can move in space, sense, and actuate through numerous channels. Thus, the total complexity of the system becomes much higher than the sum of its constituent parts.

The self-playing guitars provide a compelling way to examine how inverse interaction unsettles familiar bodily habits and reconfigures traditional ideas of performance and control. They create ground for exploring the shifting boundaries between performer and listener, and for investigating how people interpret musical intention and agency within a distributed, autonomous system.

4.4 Muzziball

Muzziball is an embodied and embedded system using both rule-based and learning-based AI to create multisensory experiences (Figure 7). It is built from a 3D-printed shell housing a Raspberry Pi, a SenseHAT shield, a battery pack, lights, and a speaker. As a rule-based system, it relies on predefined rules and conditions to trigger specific responses in its sensory and interactive components. For example, the device uses an accelerometer to detect shifts in orientation, velocity, acceleration, and jerk, responding with programmed behaviors such as changing LED colors or modifying sound output based on sensor data.

We have also explored learning-based approaches using a simple neural network running in Pure Data on the device. A sensor fusion algorithm combines data from the accelerometer, gyroscope, and magnetometer to calculate the real-time orientation of the ball. This data is fed into a trained model, which recognizes four specific orientations mapped to one of the four presets. The model, structured with two input features (pitch and roll),

---

3   A music video of the piece that combines the self-playing guitars with a learning-based system can be accessed at https://youtu.be/VmV541tXFvs.

FIGURE 7
Muzziball with interactive lights and sound activated by movement parameters.

TABLE 1 A table showing which categories the case studies fall under.

Case Study	Inverse mapping	Indirect control	Rule-based	Learning-based
Birdbox		✓	✓	
Reactive painting	✓	✓	✓	
Self-playing guitars	✓	✓	✓	✓
Muzziball		✓	✓	✓

5.1 Comparing devices

Table 1 summarizes the control types, mapping strategies, and AI types for the four systems presented above. They are all designed to respond to various actions and environmental changes, many of which have no direct connection to intentional musical expression. This opens up fascinating opportunities for incorporating sound generation into ordinary indoor spaces, such as offices, meeting rooms, and other public areas, with passive to highly interactive experiences.
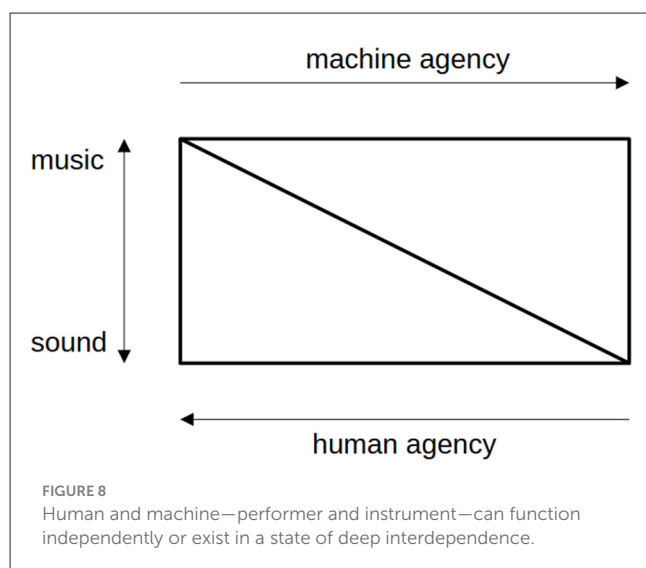
Embodiment is important in how these interactions are perceived and experienced. Perception, action, and cognition are grounded in the physical presence and movement of the body within a particular environment (Dourish, 2001). With traditional instruments, embodiment is obvious: you press keys, pluck strings, or strike surfaces to produce sound. However, musicking technologies designed for non-tactile interaction expand this idea to include indirect interactions using proximity, movement, and even environmental fluctuations that were not intended as musical gestures.

This expanded idea of embodiment is especially relevant to the concept of passive musicking—when sound generation happens without intentional interaction. Consider, for instance, an interactive Muzziball passively listening to a room through its microphone, picking up subtle sounds or pressure shifts in the air. Even when users are not deliberately engaging with it, their physical presence and unintentional movements still contribute to the evolving soundscape. Rather than reacting to clear, goal-directed gestures, such systems respond to a broader set of environmental inputs.

5.2 Agency

In this context, we use the term agency to describe the extent to which users attribute intention or intelligence to a system's behavior. Users' perception of agency often depends on how clearly they can trace the relationship between their actions and the system's output (van der Wel et al., 2012). Systems built on indirect or inverse interactions—where the connection between input and output is unclear or deliberately obscured—can end up feeling more alive or autonomous than they are. The ambiguity in these interactions can heighten engagement by making the system appear creatively responsive and adding a sense of unpredictability.

Figure 8 shows how interactive objects with embodied AI can be categorized using the varying combinations of machine and

two hidden layers with 64 neurons each, and four outputs, adjusts its output in real-time based on motion. This creates a more interactive experience, allowing sound and light to evolve in response to a user's motion. Thus, the Muzziball serves as an example of sensory augmentation (Spence and Di Stefano, 2025), where stimulation is intentionally added through a different sensory modality.

Beyond direct interaction, the Muzziball is a perpetually active system, even when left untouched in a space. Its LED array can rhythmically shift colors and intensities, creating an ambient presence that blends into its environment. We are currently exploring how active the device can be in an environment before users engage with it. Continuous, self-sustaining activity invites users; its glowing patterns serve as a signal, encouraging users to pick it up and explore its interactive capabilities. The Muzziball offers the opportunity to investigate the psychological mechanisms that shape an experience from passive to active musicking. It provides a way to study how the perception of agency changes through exploratory actions and how users attribute intentionality—and even creativity—to objects through evolving, multimodal feedback.

5 Discussion

Embodied AI introduces new possibilities for interaction, yet most contemporary AI remains disembodied, functioning through abstract computation rather than direct engagement with the physical world.

FIGURE 8
Human and machine—performer and instrument—can function independently or exist in a state of deep interdependence.

human agency they contain. Some devices are made to add musical sounds to an indoor environment (self-playing guitars), while others add nature sounds (birdbox). In these interactions, the performer and the device may act separately or participate in a more interconnected relationship where both human and machine contribute to the experience.

Ambient interaction systems are often designed to draw users' attention from the periphery toward more focused engagement, allowing them to move naturally from passive to active musicking (Tanaka, 2006). This transition from passive to active engagement occurs when users begin experimenting with their environment to influence the system's behavior. For example, subtle audio cues produced by the system might prompt someone to adjust their position in the room or intentionally create sounds to observe how the system responds. This gradual progression shows how agency and embodiment are deeply intertwined: users' bodily presence becomes a means of interacting with the system, while their perception of the system's agency depends on how well they can understand and predict the system's responses.

5.3 Intentionality

The designer or maker of an interactive system encodes particular intentions into the object, whether through programmed behaviors, sensors, or interactivity meant to provoke engagement. However, once the system is placed within an environment, the user's interpretation of that system's agency often diverges from the designer's intentions. This gap between design intention and user perception is where much of the creative potential of musicking technologies emerges. This is particularly true when interactions are indirect, inverse, or intentionally ambiguous. For example, Muzziball passively detects ambient sounds, yet users may interpret its responses as if it were purposefully reacting to their presence, even when their input is incidental or unconscious. The self-playing guitars similarly challenge conventional understandings of agency by generating sound based on environmental sonic cues, leading users to perceive these instruments as possessing a kind of

musical intentionality that operates independently of direct human control. The reactive painting further complicates this dynamic by transforming a person's proximity and motion into sound, creating an experience that feels responsive and intentional, even when the system's underlying mechanisms are relatively simple.

This perceptual agency is not static; it fluctuates based on how the user's actions—whether passive, active, or exploratory—affect the system's behavior. The ambiguity in these interactions often leads users to ascribe intentionality and even creativity to the object itself, showing how agency is constructed through the interplay between device, user, and maker. The real challenge is to design something that encourages users to keep experimenting. Figuring out the balance between passive and active modes of interaction is key to making sound a well-integrated part of everyday environments.

5.4 Embodied AI perspective

This paper reconceptualizes embodied AI by shifting focus away from models predicated on explicit, goal-directed interaction. Rather than foregrounding systems designed for deliberate control and clear causal relationships, it advances a perspective centered on subtle, environmentally embedded engagements that operate through indirect and inverse mappings. This orientation demonstrates how computationally modest systems can elicit rich, musically meaningful experiences from casual, unintentional, or even involuntary human actions.

The proposed framework characterizes embodied AI through its capacity to facilitate a multisensory engagement that is often unconscious and mediated by environmental factors. This reformulation treats the AI system as an active mediator rather than a passive tool. Designers establish direct or indirect control to allow user behaviors to influence the system, often sensed through environmental changes. Inverse mappings deliberately subvert input–output relationships, encouraging exploration and reinterpreting the conventional roles of performer and instrument. These occurrences take place at the border of awareness.

The case studies collectively show that embodied interaction need not be intentionally initiated. Instead, they engage users through presence, motion, and environmental context, giving rise to what is termed as "passive musicking"—an aesthetic experience that evolves through unconscious participation. The ambiguity and openness of these interactions often enhance their perceived agency and creativity, even when the underlying mechanisms are simple.

6 Conclusion

In this paper, we question the dominant approach to embodied AI, which often assumes direct and intentional interaction via sensing and actuation. We suggest an alternative approach based on indirect and inverse mappings. Examples of sonic and musical devices embedded in everyday environments were examined to ground this idea. While the four presented cases are centered on sonic reaction or interaction, the underlying design principles extend beyond sound, offering a general framework for embodied AI. Indirect and inverse interactions challenge the assumption that interactive systems must always demand attention, opening a space

for more fluid, background-oriented engagement. This approach raises fundamental questions: Why prioritize embodied AI? How do interactive systems navigate the space between voluntary and involuntary action or between conscious and unconscious perception? When does an interactive object shift from being part of the background to something that occupies the foreground?

Understanding these thresholds is crucial for designing AI-driven objects that complement environments without overtly capturing focus. The intention is not to create systems that compete for attention, but rather to explore how objects—whether a guitar on a stand, a painting on a wall, or interactive devices like the Muzziball and birdbox—can introduce subtle artistic interventions in everyday spaces. There is precedent for this approach in the history of sound and music: Erik Satie's furniture music and Brian Eno's ambient compositions sought to shape spaces while remaining unobtrusive (Ciaramitaro et al., 2008; Eno et al., 1978).

This kind of design work, however, also raises ethical questions. Background interactions have long been employed in commercial settings, such as shopping malls, restaurants, and retail environments, to subtly influence behavior, often in ways that serve corporate interests rather than individual experiences. This study adopts a different perspective, seeking to understand these dynamics for artistic and experiential purposes rather than manipulation. To do so requires a deeper inquiry into how embodied AI functions at the periphery of attention, influencing environments in ways that are felt rather than explicitly noticed. By refining these ideas, future research can develop interactive systems that work with, rather than against, the rhythms of everyday life.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/fourMs/self-playing-guitars.

Author contributions

MR: Writing – original draft, Conceptualization, Software. CE: Writing – review & editing. AJ: Conceptualization, Writing – review & editing.

Funding

Conflict of interest

The authors declare no affiliation with Relaxound GmbH, the company behind the Birdbox. However, this research includes a collaboration with the startup company developing the Muzziball.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. The author(s) used an AI tool for the initial generation of the teaser image; however, the image was subsequently modified and edited by the author(s) to better suit the manuscript's content. The final teaser image is not solely AI-generated but a product of both AI assistance and human refinement.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abrahamson, D., and Lindgren, R. (2014). *Embodiment and Embodied Design*. Cambridge: Cambridge University Press.

Ahmed, S. U. (2018). "Interaction and interactivity: in the context of digital interactive art installation," in *Human-Computer Interaction. Interaction in Context* (Cham: Springer), 241–257.

Barr, A., Feigenbaum, E. A., and Cohen, P. R. (1981). *The Handbook of Artificial Intelligence, Volume 3*. Austell, GA: HeurisTech Press.

Beauchamp-Williamson, L. (2004). *Band-in-a-Box. The American Music Teacher*, 98. Cincinnati, OH: Music Teachers National Association.

Bishop, L., Cancino-Chacón, C., and Goebl, W. (2019). Moving to communicate, moving to interact: Patterns of body motion in musical duo performance. *Music Percept.: an Interdiscipl. J.* 37, 1–25. doi: 10.1525/mp.2019.37.1.1

Bouchard, K., Gaboury, S., Bouchard, B., Bouzouane, A., and Giroux, S. (2016). "Smart homes in the era of big data," in *Trends in Ambient Intelligent Systems*, eds. K.

Ravulakollu, M. A. Khan, A. and Abraham (Cham: Springer International Publishing), 117–137.

Caramiaux, B., Bevilacqua, F., and Schnell, N. (2010). "Towards a gesture-sound cross-modal analysis," in *Gesture in Embodied Communication and Human-Computer Interaction*, eds. D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, et al. (Berlin: Springer Berlin Heidelberg), 158–170.

Chion, M. (1994). *Audio-vision: Sound on Screen*. New York City: Columbia University Press.

Ciaramitaro, M., Giacconi, R., Marzin, G., Ziovo, D., Hase, B., and Venice, I. (2008). *Furniture Music*. Maspalomas: Blauer Hase.

Cope, D. (2000). *The Algorithmic Composer, Volume 16*. Aiddleton, Wisconsin: AR Editions, Inc.

Dell'Anna, A., Leman, M., and Berti, A. (2021). Musical interaction reveals music as embodied language. *Front. Neurosci.* 15:667838. doi: 10.3389/fnins.2021.667838

Dourish, P. (2001). *Where the action is: The foundations of embodied interaction, volume 210*. MIT Press. doi: 10.7551/mitpress/7221.001.0001

Eno, B., Wyatt, R., Davies, R., Fast, C., Gomez, C., and Zeininger, I. (1978). "Ambient. 1. Music for airports," in *Virgin Records*. London.

Erdem, C. (2022). *Controlling or Being Controlled? Exploring Embodiment, Agency and Artificial Intelligence in Interactive Music Performance* (Doctoral thesis). Oslo: University of Oslo.

Erdem, C., Wallace, B., and Jensenius, A. R. (2022). "CAVI: a coadaptive audiovisual instrument-composition," in *NIME 2022*.

Franinovic, K., and Serafin, S., editors. (2013). Sonic interaction design. MIT Press. doi: 10.7551/mitpress/8555.001.0001

Frischmann, T. B., Mouloua, M., and Procci, K. (2015). 3-D gaming environment preferences: Inversion of the *Y* -axis. *Ergonomics* 58, 1792–1799. doi: 10.1080/00140139.2015.1044573

Ganssle, J. (2008). *The Art of Designing Embedded Systems*. Burlington, MA: Newnes.

Gaye, L., Holmquist, L. E., Behrendt, F., and Tanaka, A. (2006). "Mobile music technology: Report on an emerging community," in *NIME'06: Proceedings of the 2006 conference on New Interfaces for Musical Expression* (New York, NY: ACM Press), 22–25.

Geronazzo, M., and Serafin, S. (2023). "Sonic interactions in virtual environments," in *Human–Computer Interaction Series* (Cham: Springer International Publishing).

Gibson, J. J. (1977). "The theory of affordances," in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, eds. R. Shaw, and J. Bransford (Hillsdale, NJ: Erlbaum), 67–82.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Godøy, R. I., Haga, E., and Jensenius, A. R. (2006). "Playing "air instruments": mimicry of sound-producing gestures by novices and experts," in *Gesture in Human-Computer Interaction and Simulation*, eds. S. Gibet, N. Courty, and J. F. Kamp (Berlin Heidelberg: Springer), 256–267.

Gonzalez Sanchez, V. E., Martin, C. P., Zelechowska, A., Bjerkestrand, K. A. V., Johnson, V., and Jensenius, A. R. (2018). "Bela-based augmented acoustic guitars for sonic microinteraction," in *Proceedings of the International Conference on New Interfaces for Musical Expression* (Blacksburg, VA: Virginia Tech), 324–327.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: The MIT Press.

Groover, M. P. (2016). *Automation, Production Systems, and Computer-Integrated Manufacturing*. Delhi: Pearson Education India.

Haswell-Martin, R., Upham, F., Høffding, S., and Nielsen, N. (2025). Embodied, exploratory listening in the concert hall. *Behav. Sci*. 15:710. doi: 10.3390/bs15050710

Headlee, K., Koziupa, T., and Siwiak, D. (2010). "Sonic virtual reality game: how does your body sound?," in *Proceedings of the International Conference on New Interfaces for Musical Expression* (Princeton: Citeseer), 423–426.

Henzinger, T. A., and Sifakis, J. (2007). The discipline of embedded systems design. *Computer* 40, 32–40. doi: 10.1109/MC.2007.364

Hiller, L., and Isaacson, L. (1993). "Musical Composition with a High-Speed Digital Computer," in *Machine Models of Music*, eds. S. M. Schwanauer and D. A. Levitt (Cambridge, MA: MIT Press), 9–21.

Holland, S., Mudd, T., Wilkie-McKenna, K., McPherson, A., and Wanderley, M. M. (2019). *New Directions in Music and Human-Computer Interaction*. Cham: Springer.

Ichino, J., and Nao, H. (2018). "Playing the body: making music through various body movements," in *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion* (New York: ACM), 1–8.

Jacucci, G., Wagner, M., Wagner, I., Giaccardi, E., Annunziato, M., Breyer, N., et al. (2010). "ParticipArt: Exploring participation in interactive art installations," in *2010 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities* (Seoul: IEEE).

Jensenius, A. R. (2022). *Sound Actions: Conceptualizing Musical Instruments*. Cambridge, MA: MIT Press.

Jensenius, A. R. (2024). *Sonic Design: Explorations Between Art and Science, volume 12 of Current Research in Systematic Musicology*. Cham: Springer Nature Switzerland.

Jensenius, A. R., and Lyons, M. J. (2017). *A NIME Reader: Fifteen Years of New Interfaces*. Berlin: Springer.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM*. 60, 84–90. doi: 10.1145/3065386

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology*. Cambridge, MA: MIT Press.

Lesaffre, M., Maes, P.-J., and Leman, M. (2017). *The Routledge Companion to Embodied Music Interaction*. New York: Routledge.

Maes, P.-J., Leman, M., Palmer, C., and Wanderley, M. M. (2014). Action-based effects on music perception. *Front.Psychol*. 4:1008. doi: 10.3389/fpsyg.2013.01008

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval, Volume 39*. Cambridge: Cambridge University Press.

Martin, C. P., Jensenius, A. R., and Torresen, J. (2018). "Composing an ensemble standstill work for Myo and Bela," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, eds. L Dahl, and T. M. Douglas Bowman (Blacksburg, VI: Virginia Tech), 196–197.

Mice, L., and McPherson, A. P. (2021). "Embodied cognition in performers of large acoustic instruments as a method of designing new large digital musical instruments," in *Perception, Representations, Image, Sound, Music*, eds. R. Kronland-Martinet, S. Ystad, and M. Aramaki (Cham: Springer International Publishing), 577–590.

Miranda, E. R. (2021). *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Cham: Springer International Publishing.

Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. New York: Farrar, Straus and Giroux.

Moens, B. (2018). *D-jogger: An Interactive Music System for Gait Synchronisation with Applications for Sports and Rehabilitation* (PhD Thesis). Ghent University, Ghent, Belgium.

Morreale, F., Moro, G., Chamberlain, A., Benford, S., and McPherson, A. P. (2017). "Building a maker community around an open hardware platform," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado: ACM), 6948–6959.

Norman, D. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. New York: Basic Books.

Nymoen, K., Chandra, A., Glette, K., and Torresen, J. (2014). "Decentralized harmonic synchronization in mobile music systems," in *2014 IEEE 6th International Conference on Awareness Science and Technology (iCAST)* (Paris: IEEE), 1–6.

Owen, S. L. (2015). *Student Perceptions of the Efficacy of SmartMusic Practice Software*. Long Beach: California State University.

Pauletto, S. (2024). "Sonification and sustainability," in *The Routledge Handbook of Sound Design* (New York: Focal Press), 304–317.

Riaz, M., Guo, J., Serdar Göksülük, B., and Jensenius, A. R. (2025a). "Where is that bird? The impact of artificial birdsong in public indoor environments," in *Proceedings of the 20th International Audio Mostly Conference* (New York: ACM).

Riaz, M., Theodoridis, I., Erdem, C., and Jensenius, A. R. (2025b). "VentHackz: exploring the musicality of ventilation systems," in *Proceedings of the International Conference on New Interfaces for Musical Expression*.

Rowe, R. (1992). *Interactive Music Systems: Machine Listening and Composing*. Cambridge, MA: MIT Press.

Russell, S. J., and Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Bengaluru: Pearson.

Sandred, O. (2021). "Constraint-solving systems in music creation," in *Handbook of Artificial Intelligence for Music* (Cham: Springer), 327–344.

Schaeffer, P. (1967). *Solfège de l'objet sonore*. Paris: INA/GRM.

Schutz, M., and Lipscomb, S. (2007). Hearing gestures, seeing music: vision influences perceived tone duration. *Perception* 36, 888–897. doi: 10.1068/p5635

Shneiderman, B., and Plaisant, C. (2010). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Bengaluru: Pearson Education India.

Shortliffe, E. (2012). *Computer-Based Medical Consultations: MYCIN, volume 2*. London: Elsevier.

Small, C. (1998). *Musicking: The Meanings of Performing and Listening*. Middletown: Wesleyan University Press.

Spence, C., and Di Stefano, N. (2024). Sensory translation between audition and vision. *Psychonomic Bullet. Rev*. 31, 599–626. doi: 10.3758/s13423-023-02343-w

Spence, C., and Di Stefano, N. (2025). Augmenting art crossmodally: possibilities and pitfalls. *Front. Psychol*. 16:1605110. doi: 10.3389/fpsyg.2025.1605110

Steinbeck, P. (2018). "George Lewis's voyager," in *The Routledge Companion to Jazz Studies* (London: Routledge), 261–270.

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement learning: An Introduction*. Cambridge, MA: The MIT Press.

Tanaka, A. (2006). "Interaction, experience and the future of music," in *Consuming Music Together*, eds. K. O'Hara, and B. Brown (Berlin/Heidelberg: Springer-Verlag), 267–288.

Van Den Stock, J., Peretz, I., Grèzes, J., and De Gelder, B. (2009). Instrumental music influences recognition of emotional body language. *Brain Topogr*. 21, 216–220. doi: 10.1007/s10548-009-0099-0

van der Wel, R. P. R. D., Sebanz, N., and Knoblich, G. (2012). "Action perception from a common coding perspective," in *People Watching*, eds. K. Johnson, and M. Shiffrar (Oxford: Oxford University Press), 101–118.

Vigliensoni, G., and Fiebrink, R. (2025). Data- and interaction-driven approaches for sustained musical practices with machine learning. *J. New Music Res.* 2025, 1–14. doi: 10.1080/09298215.2024.2442361

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition* 101, 80–113. doi: 10.1016/j.cognition.2005.09.003

Wessel, D., and Wright, M. (2002). Problems and prospects for intimate musical control of computers. *Comp. Music J.* 26, 11–22. doi: 10.1162/014892602320582945

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bullet. Rev.* 9, 625–636. doi: 10.3758/BF03196322