



OPEN ACCESS

EDITED BY

Roger K Moore,
The University of Sheffield, United Kingdom

REVIEWED BY

Erik Lagerstedt,
University of Gothenburg, Sweden

*CORRESPONDENCE

Adriana Hanulíková,
✉ hanulikova@idf.uni-heidelberg.de

RECEIVED 15 October 2025

REVISED 16 December 2025

ACCEPTED 22 December 2025

PUBLISHED 12 January 2026

CITATION

Hanulíková A, Tolksdorf NF and Kapp S (2026)
Robot speech: how variability matters for
child–robot interactions.
Front. Robot. AI 12:1725423.
doi: 10.3389/frobt.2025.1725423

COPYRIGHT

© 2026 Hanulíková, Tolksdorf and Kapp. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Robot speech: how variability matters for child–robot interactions

Adriana Hanulíková*, Nils Frederik Tolksdorf and Sarah Kapp

Smart Cognition Lab, Institute for German as Foreign Philology, Heidelberg University, Heidelberg, Germany

Spoken language is one of the most powerful tools for humans to learn, exchange information, and build social relationships. An inherent feature of spoken language is large within- and between-speaker variation across linguistic levels, from sound acoustics to prosodic, lexical, syntactic, and pragmatic choices that differ from written language. Despite advancements in text-to-speech and language models used in social robots, synthetic speech lacks human-like variability. This limitation is especially critical in interactions with children, whose developmental needs require adaptive speech input and ethically responsible design. In child–robot interaction research, robot speech design has received less attention than appearance or multimodal features. We argue that speech variability in robots needs closer examination, considering both how humans adapt to robot speech and how robots could adjust to human speech. We discuss three tensions: (1) feasibility, because dynamic human speech variability is technically challenging to model; (2) desirability, because variability may both enhance and hinder learning, usability, and trust; and (3) ethics, because digital human-like speech risks deception, while robot speech varieties may support transparency. We suggest approaching variability as a design tool while being transparent about the robot's role and capabilities. The key question is which types of variation benefit children's socio-cognitive and language learning, at which developmental stage, in which context, depending on the robot's role and persona. Integrating insights across disciplines, we outline directions for studying how specific dimensions of variability affect comprehension, engagement, language learning, and for developing vocal interactivity that is engaging, ethically transparent, and developmentally appropriate.

KEYWORDS

child-robot interactions, robot speech, robot variety, robot voice, robot-directed speech, speaking style, speech variation, variability

1 Introduction

Social robots are a promising educational technology for children to support personalized and engaging learning (Peter and van Straten, 2024). However, it remains unclear how these robots should speak and sound. Spoken interaction is central to human communication and learning. In this paper, we define learning broadly as adaptive changes in children's linguistic, cognitive, and socio-emotional skills. Because speech is both the medium and the content of much early learning, the way robots speak can directly influence how children interpret,

imitate, and reason about communication itself. However, partly due to technological challenges in child–robot interaction (CRI) research, robot speech design and speech variability have received considerably less attention than other features such as appearance (Chien et al., 2025), role assignment (Rohlfing et al., 2022), or multimodal behaviors including gaze (Admoni and Scassellati, 2017) or gestures (Vogt et al., 2019; de Wit et al., 2020).

Rapid developments in artificial intelligence (AI) enable social robotics to move beyond Wizard-of-Oz (i.e., teleoperated) paradigms toward more naturalistic dialogs, using autonomously acting social robots connected to large language models (LLMs) and text-to-speech synthesis (TTS; Maure and Bruno, 2025). Synthetic speech is optimized for human-like intelligibility but usually shows less prosodic and acoustic variability than natural speech (Galdino et al., 2024). Social robots equipped with LLMs enable children to interact and learn through synthesized speech in the physical world instead of a virtual space. This disconnect between natural and synthetic speech generated by TTS technologies and LLMs raises a key design problem. Which kinds of speech variability support children's learning across developmental stages and varying language proficiencies, in which contexts, and how can technically standardized voices be balanced with pedagogical and ethical needs for natural variation?

This perspective paper advocates a child-centered and bidirectional approach to robot speech design. Robot speech should exhibit variability but not mimic human speech in all its facets. It should be designed to adapt to children's communicative requirements, contexts, and preferences. In what follows, we discuss insights from speech perception research, developmental psychology, and CRI studies to propose directions for studying how vocal design and persona, that is the robot's designed social and communicative character, can support learning, trust, and developmental appropriateness in CRI.

2 Human speech variability: functions and implications for children's learning

To understand how robot speech could or should vary, it is necessary to first consider the functions of the remarkable variability in human speech (e.g., Hawkins, 2003). Across and within speakers we encounter varying speaking rates, different levels of formality, reductions, disfluencies, repairs, diverse accents and dialects, and rich prosodic patterns. Speech also conveys social and indexical meaning such as speaker identity, age, emotions, socioeconomic and cultural background (e.g., Eckert, 2019). Speech variability is particularly relevant for children's ability to robustly discriminate speech sounds, segment words, acquire vocabulary and grammar, master pronunciation (e.g., Cristia, 2013; Rowe, 2008), as well as to learn pragmatic skills such as turn-taking and conversational repair.

For developing learners, this variability strengthens linguistic representations and supports generalization. Studies show that input variability influences the acquisition of phonology (Lively et al., 1993; Sadakata and McQueen, 2013; Hanulíková, 2023), vocabulary (Barcroft and Sommers, 2005; Levy and Hanulíková, 2023), and morphosyntax (Eidsvåg et al., 2015; Gómez, 2002). Exposure to diverse language varieties provides

redundant acoustic information that supports robust word recognition and facilitates generalization to new contexts (e.g., Potter and Saffran, 2017; Hanulíková and Ekström, 2017; Levy et al., 2019; Hanulíková and Levy, 2025).

To ground the discussion in concrete empirical examples, Table 1 maps speech variability dimensions to learning functions, risks, and design implications for CRI. The rows address different linguistic levels (phonetic/phonological, lexical/(morpho)syntactic, pragmatic, and discourse) where variability has multiple forms or functions and specifies its developmental relevance. For example, phonetic variability (row 1) supports attention and word segmentation (e.g., Cristia, 2013; Potter and Saffran, 2017), lexical diversity (row 2) enhances semantic networks (e.g., Hadley et al., 2019), while pragmatic features (row 3) like disfluencies can contribute to engagement but risk over-attribution of competence (e.g., Wigdor et al., 2016).

Speech variability serves not only learning but also social-cognitive functions. It conveys socio-indexical information that signals intentions, emotional states, and group affiliations (e.g., Kinzler, 2021). Children use these cues to guide social decisions, such as preferring speakers of their own language variety or local accent over speakers of other varieties (Kinzler, 2021), a pattern observed across developmental stages and bilingual contexts (Byers-Heinlein et al., 2017; Hanulíková, 2024).

These developmental characteristics distinguish children from adult robot users: ongoing language acquisition makes input quality critical, developing social cognition makes children particularly responsive to vocal cues about trustworthiness and competence, and their limited understanding of artificial agents makes transparency essential (Sharkey and Sharkey, 2021). Seen through this lens, the dimensions summarized in Table 1 reflect not fixed design recommendations but developmentally contingent resources.

Importantly, the effectiveness of variability depends on its type, source, timing, and relevance to the learning task (Raviv et al., 2022). Rost and McMurray (2010) found that variability along linguistically irrelevant dimension (e.g., speaker voice characteristics) helps learners identify the dimensions of input to attend to, as opposed to those they can ignore. Moreover, the type of "useful" variability may depend on the stage of learning, so that in the very early stages, variability along linguistically irrelevant dimensions is most beneficial, while later, variability along relevant dimensions become more useful (Lev-Ari, 2018). A well-known example is child-directed speech (CDS), where adults and older children systematically adjust prosody, speaking rate, lexical and syntactic complexity when speaking to toddlers (Rowe, 2008; Cristia, 2013; Kempe et al., 2024). CDS demonstrates that adaptive variability serves as an implicit scaffold for learning, supporting attention (Soderstrom, 2007), emotion regulation (Singh et al., 2002) and language development (Rowe, 2008). In designing robot speech, a similar principle could guide synthetic voices toward functional rather than fully human-like variability.

Given these learning benefits, it is unclear whether synthetic speech in social robots can become its own communicative variety, a *robot variety*, that incorporates functional variability, developmental tuning, while remaining transparent about its artificial nature. Current implementations of robot speech are relatively uniform and barely dynamically adjust to a child's abilities (Romeo et al., 2025; Kory-Westlund and Breazeal, 2019;

TABLE 1 Summary of possible key design dimensions for speech variability in child–robot interaction, integrating functional and formal perspectives. The relevance and suitability of these dimensions will vary with children’s age, learning stage, and interactional context.

Linguistic dimension	Examples of variability	Learning-relevant functions for children	Risks/challenges	CRI speech design implications
Phonetic/phonological	Pitch, rhythm, stress, timing, acoustics of sounds, pauses, disfluencies, hesitation markers, voice, accents, dialects	Increases attention and motivation, supports word segmentation, scaffolds phonological representations, adaptation and generalization, improves turn-taking, supports social inclusivity through exposure to diverse speech patterns	Overstimulation, confusion if inconsistent, timing errors disrupt interaction, limited TTS control	Adaptive modulation by task and age, include modeled disfluencies, adjustable latency
Lexical/(morpho)syntactic	Lexical diversity, gender and case marking, contextual reduction, sentence and syntactic complexity	Enhances generalization, strengthens semantic networks, supports generalization to novel forms	May hinder comprehension for low-proficiency learners, uneven benefits across developmental stages	Scaffold input from simplified to varied and more complex linguistic forms
Pragmatic/discourse	Style, register, expressiveness, pauses, disfluencies, hesitation markers, turn-taking, cultural variation (e.g., repair strategies, politeness conventions, backchannel rate)	Supports engagement, trust, and role understanding, improves conversational naturalness, supports conversational grounding across speakers and contexts	Over-attribution of competence, risk of deception if too human-like, risk of over- or under-representing certain communicative styles, leading to social or cultural bias. Recognition errors	Maintain transparency about artificial agency, use of mechanomorphic but expressive voices, curate balanced, diverse speech databases, context-sensitive pragmatics

Rohlfing et al., 2022). Thus, the question is whether the benefits of natural variability reported in speech perception research can inform the design of robot speech. Moreover, the design of robot speech must consider not only acoustic and prosodic variability but also the alignment of verbal and nonverbal behaviors, particularly as these factors interact with children’s developmental stages and interactional contexts (Wróbel et al., 2023).

3 Robot speech in CRI

3.1 Current technological implementations and challenges

Recent progress in neural TTS synthesis has made it possible to generate speech and voices that sound natural to humans (Le Maguer and Cowan, 2021). However, these systems have several limitations. Most are trained on monologic data such as audiobooks, which poorly match conversational speech (Moore, 2019). As a result, they struggle to reproduce conversational dynamics, disfluencies, and prosodic variability characteristic of spontaneous interactions (Moore and Nicolao, 2017; Moore, 2020). Temporal features such as response latency, pause placement, and overlap timing are not merely technical constraints but meaningful interactional signals that children use to infer understanding and agency. Moreover, existing models tend to suppress natural variation in speaking style, accent, dialect, register, and persona (Le Maguer and Cowan, 2021; Moore, 2019). Field studies confirm that accent mismatches can disrupt child–robot interaction and hinder learning outcomes (Singh et al., 2023). Although work on incremental language generation and speech synthesis creates more natural

turn-taking through pauses, repetitions and repair (Buschmeier and Kopp, 2018), most current systems still lack the dynamic variability that characterizes human communication (Ekstedt and Skantze, 2022).

A further challenge concerns child-specific interaction. Automatic speech recognition systems still struggle with children’s variable or developmentally atypical speech patterns (Kennedy et al., 2017; Janssens et al., 2025), and generative models lack sufficient child-directed training data. Developing more ecologically valid models requires high-quality, annotated interactional datasets that capture real-world dialogic dynamics. While recent advances in child speech recognition are promising (Janssens et al., 2025), substantial data scarcity representing diverse children persists.

These technical constraints not only affect recognition accuracy but also influence how spoken interactions unfold in practical settings. Following Moore (2019), the mismatch between human-like speech and limited linguistic and interactional abilities of robots creates what he calls the “habitability gap”. If a robot sounds too human-like, children and adults may overestimate its abilities and understanding, leading to disappointment when expectations are not met. This gap is particularly problematic in CRI, where unnatural dialog patterns reduce ecological validity compared to children’s everyday language experiences, which involve highly dynamic, multimodal, and interactionally contingent input (Goldenberg et al., 2022). Studies addressing these limitations show that strategically implemented disfluencies and conversational fillers can improve turn-taking dynamics and social engagement in CRI (Ohshima et al., 2015; Wigdor et al., 2016). Nevertheless, most current systems lack the ability to adapt speech rate dynamically to children to facilitate comprehension, and they lack “priors”,

that is, built-in understanding of how language works in human communicative contexts (Moore, 2005).

Because of these challenges, some researchers suggest the use of synthetic speech as a distinct, purpose-built variety. Le Maguer and Cowan (2021) argue for “natural non-human-like speech synthesis”, i.e., voices that are intelligible and expressive but transparently artificial, while Moore (2017) proposes “mechanomorphic” designs emphasizing congruency between voice and robot’s non-human identity. Marge et al. (2022) highlight this approach to align appearance, capabilities, and voice. Moreover, their work identifies the need for interaction styles to be deliberately engineered and tuned for specific scenarios, emphasizing the role of prosody in turn-taking, grounding, or conveying stance. Moore (2017) frames this as developing a “science of vocal interactivity”, referring to a systematic investigation of how vocal design in embodied agents affects learning, trust, and social dynamics. These perspectives mainly address voice design, and the focus lies primarily on general robot users. Our proposal extends this discussion in two ways: First, we apply these ideas to developmentally appropriate robot speech varieties for children in interactional and educational settings. Second, we distinguish between voice (timbre, pitch) and speech characteristics (speech rate, prosodic and articulatory variability). For interactions with children, the challenge extends beyond naturalness. Developmentally appropriate robot speech must balance familiarity with transparency: voices that sound too human-like risk eliciting misplaced trust or over-attribution of understanding, whereas overly mechanical voices can reduce engagement and warmth. An alternative is a synthetic yet expressive voice that may best support learning, engagement and trust while signaling artificiality. Such “mechanomorphic” or hybrid voices could adapt prosodic range, rhythm, and affect to the communicative context without simulating a specific human identity. Addressing this challenge calls for a systematic investigation of both dimensions (voice and speech variability) and their interaction.

3.2 Robot speech effects on children

Research examining how robot speech characteristics affect children shows that expressive speech enhances engagement and learning. Preschoolers interacting with robots using expressive speech showed improved word production, better narrative recall, and greater engagement in storytelling tasks compared to those interacting in monotone speech (Kory-Westlund et al., 2017; Conti et al., 2019). Similar effects have also been observed in young adults, with L2 learners performing better in a linguistic task when a robot delivers instructions in a charismatic speaking style (Fischer et al., 2021). In addition, adaptive features such as entrainment, where the robot adjusts pitch, rate, and volume to match a child’s speaking style, can support rapport and positive emotions during interaction (Kory-Westlund and Breazeal, 2019). However, when robots sound more human-like, children show greater compliance with their requests (Romeo et al., 2025), raising ethical questions about transparency and possible manipulation.

Children sometimes prefer robots that make systematic errors, which create opportunities for correction and scaffolding (Förster et al., 2023). After interacting with a robot having

pronunciation difficulties, preschoolers engaged in metatalk about the robot’s voice and limitations, demonstrating emerging critical technological thinking (Tolksdorf et al., 2024). Such behaviors, including repair, clarification requests, and delayed responses constitute a critical dimension of speech variability that shape how children interpret competence, intentionality, and transparency in interaction. This suggests that strategic imperfection can support both learning and increase awareness of technology.

Studies on robot-assisted language learning show mixed outcomes. While some studies report improvements in pronunciation, vocabulary, and communicative ability (Lee et al., 2011; Wang et al., 2013), others find effects limited to listening improvements (In and Han, 2015). These inconsistencies likely reflect both varied assessment methods and the limited natural variation in TTS systems, which constrains intonation and reduces alignment opportunities for learners (Rosenthal-von der Pütten et al., 2016).

Thus, expressive, contingent, and socially responsive robot speech can promote engagement and learning, though effects remain context-dependent and methodologically fragmented.

3.3 Bidirectional adaptation and individual differences

Speech adaptation is central to CRI and learning because it both reflects interlocutors’ expectations about the robot’s communicative behavior and directly affects the structure of the input available for learning. Thus, child-robot interaction needs to be considered as a bidirectional process. Both adults and children modify their speech when addressing robots, using robot-directed speech (RDS). RDS is characterized by features such as slower rate, repetitions, simple sentence structure, and increased pitch variation (e.g., Breazeal, 2002; Cohn et al., 2021; Cohn et al., 2024), features similar to CDS. These modifications reflect assumptions about the robot’s cognitive and linguistic capabilities (Fischer et al., 2011) and its perceived competence rather than anthropomorphism: less capable robots elicit stronger speech adjustments (Cohn et al., 2024; Kalashnikova et al., 2023). Adults show prosodic and phonetic alignment with synthetic voices (Zellou et al., 2021; Offrede et al., 2023) and even converge on emotional expressiveness (Cohn et al., 2021).

Children also show phonetic accommodation to robot speech, adjusting fundamental frequency, vowel duration, and vowel quality, with considerable individual variation predicted by personality traits and perception of the robot’s persona (Hong and Chen, 2024). Cohn et al. (2024) showed that children’s vocal modifications are more extreme than adults’, demonstrating that age is a critical factor in RDS. Such adaptation occurs at implicit, cortical levels. Sivridag and Mani (2025) found that 5-year-olds’ brains tracked both synthesized robot speech and natural adult speech, though with longer processing delays for robot speech, indicating cortical entrainment during child-robot interaction.

Interestingly, children’s RDS may reflect interpersonal dynamics. Verner et al. (2024) found that children’s vocal characteristics (pitch variation, intensity) correlated with their trust in the robot, though effects were small. Sanoubari et al. (2024) demonstrated that prosody can disambiguate spoken input during human-robot interaction,

with participants using distinct prosodic patterns to convey different intentions with identical words (e.g., “nice” meaning “keep going” vs. “stop”). This suggests that the design of robot speech and the capability of systems to reliably perceive children’s speech patterns must account for how children adapt to robots, not just how robots adapt to children.

Just as caregivers adjust their speech to individual children, robots should ideally do the same. Research on CRI has only begun to explore such adaptive patterns. While children’s vocabulary size, phonological memory, and selective attention moderate robot-assisted learning outcomes (Van den Berghe et al., 2021; Rudenko et al., 2024), few studies systematically tailor robot speech to these differences. In contrast to some work with adults (Crumpton and Bethel, 2016; Skantze et al., 2019), existing CRI learning studies (e.g., Vogt et al., 2019) tend to prioritize other communicative modalities, such as the effect of gestures. Consequently, designs that systematically adapt robot speech parameters such as speaking rate, prosody, or disfluency to individual developmental needs remain underexplored and technically limited, despite children’s vocal accommodation providing a potential signal for adaptive robot behavior.

4 Should robots embrace speech variability?

The preceding sections highlight speech variability as both an opportunity and a challenge for CRI, giving rise to three interrelated tensions. The first concerns feasibility, because modeling dynamic human speech variability remains technically challenging. While LLM-driven speech synthesis enables prosodic and persona-level variation, fine-grained phonetic timing or natural disfluencies still pose an issue.

The second tension concerns desirability, because speech variability can shape engagement, trust and learning in positive and negative ways. Variability strengthens linguistic representations when appropriately structured (Rost and McMurray, 2010), and expressive robot voices enhance engagement and language learning (Kory-Westlund et al., 2017; Conti et al., 2019). However, excessive or poorly timed variability risks confusion. The optimal degree of variability requires balance: too little sounds mechanical, too much undermines intelligibility and attention. Finally, for certain language-learning contexts, fine-grained phonetic variability may be equally important as prosodic expressiveness or persona cues, particularly because these dimensions interact in shaping how children perceive and adapt to robot speech, though systematic comparisons are lacking.

The third tension concerns ethics. When synthetic speech becomes indistinguishable from natural human speech, it risks obscuring a robot’s actual capabilities, leading to deception and over-attribution of competencies (Sharkey and Sharkey, 2021). Transparently synthetic yet engaging speech makes the robot’s capabilities and artificial nature clear while supporting meaningful interaction (Moore, 2017). Additional risks include increased emotional attachment and unintended reinforcement of stereotypes.

These tensions demonstrate that speech variability is not merely a technical or aesthetic concern but one with direct developmental, pedagogical, and ethical implications. We therefore

define robot speech varieties as systematic, adaptive synthetic voices that remain transparently artificial but incorporate functional variation to support specific learning and communicative goals. Thus, we do not refer to a stable linguistic system similar to a human dialect, but to a principled design abstraction specifying task- and age-sensitive speech parameters. Based on proposals for “natural non-human-like speech” (Le Maguer and Cowan, 2021) and “mechanomorphic” voices (Moore, 2017), an ethical design of robot speech emphasizes transparency about the robot’s technological nature while implementing variation that support children’s development and engagement and is aligned with the robot’s role. The key question shifts from whether robots should embrace human variability to which types (phonetic and phonological features, persona characteristics, temporal dynamics and their interactions) benefit which learning goals and at what stage during development. Importantly, not all variability can be implemented at the acoustic signal level, some forms are better realized through interaction management, turn-taking strategies, or repair policies, given current system limitations. Future studies should experimentally test these parameters across developmental groups.

5 Discussion and looking ahead

By integrating insights across disciplines, we suggest that targeted implementation and accommodation of speech variability can be a valuable strategy in CRI, provided it is informed by technical feasibility, cognitive and contextual demands, and ethical transparency. While prior work has addressed general human–robot interactions (Marge et al., 2022; Huang and Moore, 2025), our perspective focuses on children’s specific needs, including developmental appropriateness, child-directed variability, and age-sensitive design. Addressing these questions requires systematic research across multiple linguistic dimensions which exhibit variability (see Table 1). Across the dimensions discussed, we argue that speech variability can support learning, engagement, and socio-communicative development, but that its effectiveness depends on children’s developmental stage, interactional context, and the robot’s role, persona, and capabilities. Overgeneralizing from human speech variability risks misrepresenting robot competence; therefore, variation should remain functional, interpretable, and transparently artificial.

Future studies should examine how robot speech variability interacts with multimodal cues (e.g., gaze, gesture, timing) to shape trust and developmental outcomes, including linguistic and cultural diversity (e.g., Andrist et al., 2015), while minimizing bias and stereotypes, and utilizing cross-linguistic and longitudinal designs. Research should also extend beyond dyadic interactions to polyadic settings with multiple children, caregivers, or educators, investigating bidirectional speech adaptation in these socially shared contexts. Importantly, more research is needed to understand the conditions under which children recognize synthetic speech as artificial and how this metacognitive awareness develops with age.

From a technical feasibility perspective, advancing robot speech varieties requires generative models trained on child-directed, multimodal speech and improved child-speech recognition systems that can handle dialectal and developmentally variable input.

While neural TTS synthesis steadily improve, fully autonomous bidirectional adaptation between children and robots remains technically challenging. Moreover, adaptive control of prosody, clarity, and timing could improve accessibility for children with hearing, speech, or neurodevelopmental differences, allowing robot varieties to serve a broader range of learners and make CRI more inclusive.

Rather than aiming to make robots sound perfectly human, the future of CRI should treat developmental appropriateness as the primary criterion for vocal design. This shift reframes variability as a design resource to be deployed selectively, transparently, and in alignment with children's learning needs, rather than as a by-product of human-likeness. Meeting these challenges will require collaboration across diverse disciplines and cultural contexts.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

AH: Conceptualization, Funding acquisition, Writing – original draft, Writing – review and editing, Project administration. NT: Validation, Writing – review and editing. SK: Validation, Writing – review and editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was funded within

References

- Admoni, H., and Scassellati, B. (2017). Social eye gaze in human-robot interaction: a review. *J. Human-Robot Interact.* 6 (1), 25–63. doi:10.5898/JHRI.6.1.Admoni
- Andrist, A., Ziadee, M., Boukaram, H., Mutlu, B., and Sakr, M. (2015). "Effects of culture on the credibility of robot speech: a comparison between English and Arabic," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, Portland, OR, USA, 02-05 March 2015, 157–164. doi:10.1145/2696454.2696464
- Barcroft, J., and Sommers, M. S. (2005). Effects of acoustic variability on second language learning. *Stud. Second Lang. Acquis.* 27 (3), 387–414. doi:10.1017/S0272263105050175
- Bradlow, A. R., and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition* 106 (2), 707–729. doi:10.1016/j.cognition.2007.04.005
- Breazeal, C. (2002). Regulation and entrainment in human–robot interaction. *Int. J. Robotics Res.* 21 (10–11), 883–902. doi:10.1177/0278364902021010096
- Brooks, R., Breazeal, C., and Scassellati, B. (2010). *The expressive robot: new approaches to human-robot interaction*. MIT Press.
- Buschmeier, H., and Kopp, S. (2018). "Communicative listener feedback in human-agent interaction: artificial speakers need to be attentive and adaptive," in *Proceedings of the 17th international conference on autonomous agents and multiagent systems* (Stockholm, Sweden: ACM), 1213–1221.
- Byers-Heinlein, K., Behrend, D. A., Said, L. M., Girgis, H., and Poulin-Dubois, D. (2017). Monolingual and bilingual children's social preferences for monolingual and bilingual speakers. *Dev. Sci.* 20 (4), e12392. doi:10.1111/desc.12392
- Chien, S.-E., Chen, Y.-S., Chen, Y.-C., and Yeh, S.-L. (2025). Exploring the developmental aspects of the uncanny valley effect on children's preferences for robot appearance. *Int. J. Human-Computer Interact.* 41 (10), 6366–6376. doi:10.1080/10447318.2024.2376365
- Cohn, M., Predeck, K., Sarian, M., and Zellou, G. (2021). Prosodic alignment toward emotionally expressive speech: comparing human and alexa model talkers. *Speech Commun.* 135, 66–75. doi:10.1016/j.specom.2021.10.003
- Cohn, M., Barrera, S., Graf Estes, K., Yu, Z., and Zellou, G. (2024). Children and adults produce distinct technology- and human-directed speech. *Sci. Rep.* 14 (1), 15611. doi:10.1038/s41598-024-66313-5
- Conti, D., Cattani, A., Di Nuovo, S., and Di Nuovo, A. (2019). Are future psychologists willing to accept and use a humanoid robot in their practice? Italian and English students' perspective. *Front. Psychol.* 10, 2138. doi:10.3389/fpsyg.2019.02138
- Cristia, A. (2013). Input to language: the phonetics and perception of infant-directed speech: the phonetics and perception of infant-directed speech. *Lang. Linguistics Compass* 7 (3), 157–170. doi:10.1111/lnc3.12015
- Crumpton, J., and Bethel, C. L. (2016). A survey of using vocal prosody to convey emotion in robot speech. *Int. J. Soc. Robotics* 8, 271–285. doi:10.1007/s12369-015-0329-4
- de Wit, J., Brandse, A., Krahmer, E., and Vogt, P. (2020). "Varied human-like gestures for social robots: investigating the effects on children's engagement and language learning," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, Cambridge, MA, USA, 23-26 March 2020 (IEEE), 359–367.
- Eckert, P. (2019). The limits of meaning: social indexicality, variation, and the cline of interiority. *Language* 95 (4), 751–776. doi:10.1353/lan.0.0239

the framework of the Excellence Strategy of the Federal and State Governments, Germany (AH).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. Only for proofreading and editing, not for research, analysis, or content generation.

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.

- Eidsvåg, S. S., Austad, M., Plante, E., and Asbjørnsen, A. E. (2015). Input variability facilitates unguided subcategory learning in adults. *J. Speech, Lang. Hear. Res.* 58 (3), 826–839. doi:10.1044/2015_JSLHR-L-14-0172
- Ekstedt, E., and Skantze, G. (2022). How much does prosody help turn-taking? Investigations using voice activity projection models. arXiv:2209.05161v1.
- Fischer, K., Foth, K., Rohlfing, K., and Wrede, B. (2011). "Is talking to a simulated robot like talking to a child?" in 2011 IEEE International Conference on Development and Learning (ICDL), Frankfurt am Main, Germany, 24-27 August 2011 (IEEE), 1–6.
- Fischer, K., Niebuhr, O., and Alm, M. (2021). Robots for foreign language learning: speaking style influences student performance. *Front. Robotics AI* 8, 680509. doi:10.3389/frobt.2021.680509
- Förster, F., Romeo, M., Holthaus, P., Wood, L. J., Dondrup, C., Fischer, J. E., et al. (2023). Working with troubles and failures in conversation between humans and robots: workshop report. *Front. Robotics AI* 10, 1202306. doi:10.3389/frobt.2023.1202306
- Galdino, J. C., Araújo, G. E., Junior, A. C., Jr, M. O., Ponti, M. A., and Aluisio, S. M. (2024). Acoustic analysis of prosodic features in natural versus synthesized speech samples from YourTTS and SYNTACC models. *Encontro Nac. Inteligência Artif. Comput. (ENIAC)*, 304–315. doi:10.5753/eniac.2024.245092
- Goldenberg, E. R., Repetti, R. L., and Sandhofer, C. M. (2022). Contextual variation in language input to children: a naturalistic approach. *Dev. Psychol.* 58 (6), 1051–1065. doi:10.1037/dev0001345
- Gómez, R. (2002). Variability and detection of invariant structure. *Psychol. Sci.* 13 (5), 431–436. doi:10.1111/1467-9280.00476
- Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., and Golinkoff, R. M. (2019). Building semantic networks: the impact of a vocabulary intervention on preschoolers' depth of word knowledge. *Read. Res. Q.* 54, 41–61. doi:10.1002/rrq.225
- Hanulíková, A. (2023). "Learning phonotactically complex L3 words: are bilinguals more successful?" in 20th international congress of phonetic sciences (ICPhS) (Barcelona, Spain: ICPhS), 2701–2705.
- Hanulíková, A. (2024). Navigating accent bias in German: children's social preferences for a second-language accent over a first-language regional accent. *Front. Lang. Sci.* 3, 1357682. doi:10.3389/flang.2024.1357682
- Hanulíková, A., and Ekström, J. (2017). "Lexical adaptation to a novel accent in German: a comparison between German, Swedish, and Finnish listeners," in *Proceedings of Interspeech* (Stockholm, Sweden: Interspeech), 1784–1788. doi:10.21437/Interspeech.2017-369
- Hanulíková, A., and Levy, H. (2025). Quantifying experience with accented speech to study monolingual and bilingual school-aged children's speech processing. *Languages* 10 (4), 80. doi:10.3390/languages10040080
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phonetics* 31 (3), 373–405. doi:10.1016/j.wocn.2003.09.006
- Hong, Y., and Chen, S. (2024). "Individual variation in phonetic accommodation of Mandarin-speaking children during conversations with a virtual robot," in *Speech prosody 2024* (Leiden, The Netherlands: Leiden Universiteit), 472–476. doi:10.21437/SpeechProsody.2024-96
- Huang, G., and Moore, R. K. (2025). "Adaptive affordance design for social robots: tailoring to role-specific preferences," in 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Melbourne, Australia, 04-06 March 2025 (IEEE), 580–588.
- In, J., and Han, J. (2015). "The acoustic-phonetics change of English learners in robot assisted learning," in *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction extended abstracts* (Portland Oregon USA: ACM), 39–40. doi:10.1145/2701973.2702003
- Janssens, R., Verhelst, E., Abbo, G. A., Ren, Q., Bernal, M. J. P., and Belpaeme, T. (2025). "Child speech recognition in human-robot interaction: problem solved?," in *Social robotics*. Editors O. Palinko, L. Bodenhausen, J.-J. Cabibihan, K. Fischer, S. Šabanović, K. Winkle, et al. (Cham, Switzerland: Springer), 476–486. doi:10.1007/978-981-96-3519-1_43
- Kalashnikova, N., Hutin, M., Vasilescu, I., and Devillers, L. (2023). "The effect of human-likeness in French robot-directed speech: a study of speech rate and fluency," in *Text, speech, and dialogue*. Editors K. Ekstein, F. Pártl, and M. Konopik (Cham, Switzerland: Springer), 249–257. doi:10.1007/978-3-031-40498-6_22
- Kempe, V., Ota, M., and Schaeffler, S. (2024). Does child-directed speech facilitate language development in all domains? A study space analysis of the existing evidence. *Dev. Rev.* 72, 101121. doi:10.1016/j.dr.2024.101121
- Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., et al. (2017). "Child speech recognition in human-robot interaction: evaluations and recommendations," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, Vienna, Austria, 06-09 March 2017 (IEEE), 82–90.
- Kinzler, K. D. (2021). Language as a social cue. *Annu. Rev. Psychol.* 72, 241–264. doi:10.1146/annurev-psych-010418-103034
- Kory-Westlund, J. M., and Breazeal, C. (2019). Exploring the effects of a social robot's speech entrainment and backstory on young children's emotion, rapport, relationship, and learning. *Front. Robotics AI* 6, 54. doi:10.3389/frobt.2019.00054
- Kory-Westlund, J. M., Jeong, S., Park, H. W., Ronfard, S., Adhikari, A., Harris, P. L., et al. (2017). Flat vs. expressive storytelling: young children's learning and retention of a social robot's narrative. *Front. Hum. Neurosci.* 11, 1–20. doi:10.3389/fnhum.2017.00295
- Le Maguer, S., and Cowan, B. R. (2021). "Synthesizing a human-like voice is the easy way," in *Proceedings of the 3rd conference on conversational user interfaces* (New York, USA: ACM), 1–3. doi:10.1145/3469595.3469614
- Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. G., Sagong, S., et al. (2011). On the effectiveness of robot-assisted language learning. *ReCALL* 23, 25–58. doi:10.1017/S0958344010000273
- Lev-Ari, S. (2018). Social network size can influence linguistic malleability and the propagation of linguistic change. *Cognition* 176, 31–39. doi:10.1016/j.cognition.2018.03.003
- Levy, H., and Hanulíková, A. (2023). Spot it and learn it! word learning in virtual peer-group interactions using a novel paradigm for school-aged children. *Lang. Learn.* 73 (1), 197–230. doi:10.1111/lang.12520
- Levy, H., Konieczny, L., and Hanulíková, A. (2019). Processing of unfamiliar accents in monolingual and bilingual children: effects of type and amount of accent experience. *J. Child Lang.* 46 (2), 368–392. doi:10.1017/S030500091800051X
- Lively, S. E., Logan, J. S., and Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: the role of phonetic environment and talker variability in learning new perceptual categories. *J. Acoust. Soc. Am.* 94 (3), 1242–1255. doi:10.1121/1.408177
- Marge, M., Espy-Wilson, C., Ward, N. G., Alwan, A., Artzi, Y., Bansal, M., et al. (2022). Spoken language interaction with robots: recommendations for future research. *Comput. Speech & Lang.* 71, 101255. doi:10.1016/j.csl.2021.101255
- Maure, R., and Bruno, B. (2025). Autonomy in socially assistive robotics: a systematic review. *Front. Robotics AI* 12, 1586473. doi:10.3389/frobt.2025.1586473
- Moore, R. K. (2005). "Towards a unified theory of spoken language processing," in Fourth IEEE Conference on Cognitive Informatics (ICCI 2005), Irvine, CA, USA, 08-10 August 2005. Editor H. Huang, and M. Moore (IEEE), 167–172.
- Moore, R. K. (2017). "Appropriate voices for artefacts: some key insights," in *1st international workshop on vocal interactivity in-and-between humans, animals and robots, (Paris, France: vihar)*. Available online at: https://vihar-2017.vihar.org/assets/papers/VIHAR-2017_paper_8.pdf.
- Moore, R. K. (2019). Talking with robots: opportunities and challenges. arXiv. 10.48550/arXiv.1912.00369.
- Moore, R. K. (2020). "PCT and beyond: toward a computational framework for 'intelligent' communicative systems," in *The interdisciplinary handbook of perceptual control theory*. Editor W. Mansell (Academic Press), 557–582. doi:10.1016/B978-0-12-818948-1.00015-0
- Moore, R. K., and Nicolao, M. (2017). Toward a needs-based architecture for 'intelligent' communicative agents: speaking with intention. *Front. Robotics AI* 4, 66. doi:10.3389/frobt.2017.00066
- Offrede, T., Mishra, C., Skantze, G., Fuchs, S., and Mooshammer, C. (2023). "Do humans converge phonetically when talking to a robot?," in *Proceedings of the international congress of phonetic sciences (ICPhS)* (Prague: Guarant International), 3507–3511.
- Ohshima, N., Kimijima, K., Yamato, J., and Mukawa, N. (2015). "A conversational robot with vocal and bodily fillers for recovering from awkward silence at turn-takings," in 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Kobe, Japan, 31 August 2015 - 04 September 2015 (IEEE), 325–330.
- Peter, J., and van Straten, C. L. (2024). "Social robots and children: a field in development," in *The De Gruyter Handbook of Robots in Society and Culture*. Editors L. Fortunati, and A. Edwards (Berlin, Boston: De Gruyter), 371–388. doi:10.1515/9783110792270-020
- Potter, C. E., and Saffran, J. R. (2017). Exposure to multiple accents supports infants' understanding of novel accents. *Cognition* 166, 67–72. doi:10.1016/j.cognition.2017.05.031
- Raviv, L., Lupyan, G., and Green, S. C. (2022). How variability shapes learning and generalization. *Trends Cognitive Sci.* 26 (6), 462–483. doi:10.1016/j.tics.2022.03.007
- Rohlfing, K. J., Altvater-Mackensen, N., Caruana, N., van den Bergh, R., Bruno, B., Tolksdorf, N. F., et al. (2022). Social/dialogical roles of social robots in supporting children's learning of language and literacy—A review and analysis of innovative roles. *Front. Robotics AI* 9, 971749. doi:10.3389/frobt.2022.971749
- Romeo, M., Torre, I., Le Maguer, S., Sleat, A., Cangelosi, A., and Leite, I. (2025). The effect of voice and repair strategy on trust formation and repair in human-robot interaction. *J. Human-Robot Interact.* 14 (2), 1–22. doi:10.1145/3711938
- Rosenthal-von der Pütten, A. M., Straßmann, C., and Krämer, N. C. (2016). "Robots or agents – neither helps you more or less during second language acquisition," in *Intelligent virtual agents*. Editors D. Traum, W. Swartout, P. Khooshabeh, S. Kopp, S. Scherer, and A. Leuski (Cham, Switzerland: Springer), 256–268. doi:10.1007/978-3-319-47665-0_23

- Rost, G. C., and McMurray, B. (2010). Finding the signal by adding noise: the role of noncontrastive phonetic variability in early word learning. *Infancy* 15 (6), 608–635. doi:10.1111/j.1532-7078.2010.00033.x
- Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *J. Child Lang.* 35 (1), 185–205. doi:10.1017/S0305000907008343
- Rudenko, I., Rudenko, A., Lilienthal, A. J., Arras, K. O., and Bruno, B. (2024). The child factor in child–robot interaction: discovering the impact of developmental stage and individual characteristics. *Int. J. of Soc. Robotics* 16 (8), 1879–1900. doi:10.1007/s12369-024-01121-5
- Sadakata, M., and McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: evidence from Japanese geminates. *J. Acoust. Soc. Am.* 134 (2), 1324–1335. doi:10.1121/1.4812767
- Sanoubari, E., Iscen, A., Takayama, L., Saliceti, S., Cunningham, C., and Caluwaerts, K. (2024). Prosody for intuitive robotic interface design: it's not what you said, it's how you said it. arXiv. 10.48550/arXiv.2403.08144.
- Sharkey, A., and Sharkey, N. (2021). We need to talk about deception in social robotics. *Ethics Inf. Technol.* 23 (3), 309–316. doi:10.1007/s10676-020-09573-9
- Singh, L., Morgan, J. L., and Best, C. T. (2002). Infants' listening preferences: baby talk or happy talk? *Infancy* 3 (3), 365–394. doi:10.1207/S15327078IN0303_5
- Singh, D. K., Kumar, M., Fosch-Villaronga, E., Singh, D., and Shukla, J. (2023). Ethical considerations from child-robot interactions in under-resourced communities. *Int. J. Soc. Robotics* 15 (12), 2055–2071. doi:10.1007/s12369-022-00882-1
- Sivridag, F., and Mani, N. (2025). Children's cortical speech tracking in child–adult and child–robot interactions. *Dev. Psychol.* doi:10.1037/dev0002086
- Skantze, G., Gustafson, J., and Beskow, J. (2019). “Multimodal conversational interaction with robots,” in *The handbook of multimodal-multisensor interfaces: language processing, software, commercialization, and emerging directions - volume 3*. doi:10.1145/3233795.3233799
- Soderstrom, M. (2007). Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Dev. Rev.* 27 (4), 501–532. doi:10.1016/j.dr.2007.06.002
- Tolksdorf, N. F., Wildt, E., and Rohlfing, K. J. (2024). “Preschoolers' interactions with social robots: investigating the potential for eliciting metatalk and critical technological thinking,” in *Companion of the 2024 ACM/IEEE international conference on human-robot interaction* (Stockholm, Sweden: ACM/IEEE), 1053–1057. doi:10.1145/3610978.3640654
- van den Berghe, R., de Haas, M., Oudgenoeg-Paz, O., Kraemer, E., Verhagen, J., Vogt, P., et al. (2021). A toy or a friend? children's anthropomorphic beliefs about robots and how these relate to second-language word learning. *J. Comput. Assisted Learn.* 37 (2), 396–410. doi:10.1111/jcal.12497
- Velner, E., Beelen, T., Schadenberg, B., Ordelman, R., Huibers, T., Truong, K. P., et al. (2024). “Uhm are you sure?” an exploratory study of trust indicators in robot-directed child speech,” in *Proceedings of the ACM international conference on intelligent virtual agents* (Glasgow, United Kingdom: ACM), 1–4. doi:10.1145/3652988.3673933
- Vogt, P., Van Den Berghe, R., De Haas, M., Hoffman, L., Kanero, J., Mamus, E., et al. (2019). “Second language tutoring using social robots: a large-scale study,” in 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Daegu, Korea (South), 11–14 March 2019 (IEEE), 497–505.
- Wang, Y. H., Young, S. S.-C., and Jang, J.-S. R. (2013). Using tangible companions for enhancing learning English conversation. *J. Educ. Technol. & Soc.* 16, 296–309. doi:10.1109/ICALT.2010.190
- Wigdor, N., de Greeff, J., Looije, R., and Neerinx, M. A. (2016). “How to improve human-robot interaction with conversational fillers,” in 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016 (IEEE), 219–224.
- Wróbel, A., Żróbek, K., Schaper, M.-M., Zguda, P., and Indurkha, B. (2023). “Age-appropriate robot design: in-the-wild child-robot interaction studies of perseverance styles and robot's unexpected behavior,” in 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Kanazawa, Japan, 28–31 August 2023 (IEEE), 1451–1458.
- Zellou, G., Cohn, M., and Ferenc Segedin, B. (2021). Age- and gender-related differences in speech alignment toward humans and voice-AI. *Front. Commun.* 5, 600361. doi:10.3389/fcomm.2021.600361