



## OPEN ACCESS

## EDITED BY

Frank Foerster,  
University of Hertfordshire, United Kingdom

## REVIEWED BY

Dimosthenis Kontogiorgos,  
Royal Institute of Technology, Sweden

## \*CORRESPONDENCE

Angus Addelee, [a.addelee@hw.ac.uk](mailto:a.addelee@hw.ac.uk)

RECEIVED 15 December 2023

ACCEPTED 21 November 2024

PUBLISHED 03 January 2025

## CITATION

Addelee A and Papaioannou I (2025) Building for speech: designing the next-generation of social robots for audio interaction. *Front. Robot. AI* 11:1356477. doi: 10.3389/frobt.2024.1356477

## COPYRIGHT

© 2025 Addelee and Papaioannou. 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.

# Building for speech: designing the next-generation of social robots for audio interaction

Angus Addelee<sup>1\*</sup> and Ioannis Papaioannou<sup>2</sup>

<sup>1</sup>Department of Mathematics and Computer Science, Heriot-Watt University, Edinburgh, United Kingdom, <sup>2</sup>Aveni AI, Edinburgh, United Kingdom

There have been significant advances in robotics, conversational AI, and spoken dialogue systems (SDSs) over the past few years, but we still do not find social robots in public spaces such as train stations, shopping malls, or hospital waiting rooms. In this paper, we argue that early-stage collaboration between robot designers and SDS researchers is crucial for creating social robots that can legitimately be used in real-world environments. We draw from our experiences running experiments with social robots, and the surrounding literature, to highlight recurring issues. Robots need better speakers, a greater number of high-quality microphones, quieter motors, and quieter fans to enable human-robot spoken interaction in the wild. If a robot was designed to meet these requirements, researchers could create SDSs that are more accessible, and able to handle multi-party conversations in populated environments. Robust robot joints are also needed to limit potential harm to older adults and other more vulnerable groups. We suggest practical steps towards future real-world deployments of conversational AI systems for human-robot interaction.

## KEYWORDS

social robots, spoken dialogue, accessibility, robotics, human-robot interaction, conversational AI

## 1 Introduction

Social robots are not yet found in our public spaces, despite this vision being an imminent reality over 25 years ago (Thrun, 1998). They do not roam our shopping malls helping lost families find the bathroom, we do not bump into them providing departure times in train stations and airports, and they are not helping patients in hospital waiting rooms with their questions (see Figure 1).

Spoken dialogue systems (SDSs) have consistently improved over time (Glass, 1999; Williams, 2009; Lemon, 2022), with many years of peer-reviewed papers containing remarkable results. These models, however, are often evaluated automatically upon collected data, or with users in highly controlled lab settings. Social robots work wonderfully in the lab, but fail when deployed in the real world for experiments or demonstration (Tian and Oviatt, 2021). Some of these failures stem from the embedded SDS [for example: multi-party interactions (Addelee et al., 2023a), socio-affective competence (Tian and Oviatt, 2021), voice accessibility (Addelee, 2023), or trust failures (Tolmeijer et al., 2020)], but even interactions that the SDS should be able to handle with ease go wrong. These failures are often caused by the design of the social robot itself, classified by Honig and Oron-Gilad (2018) as technical hardware failures.

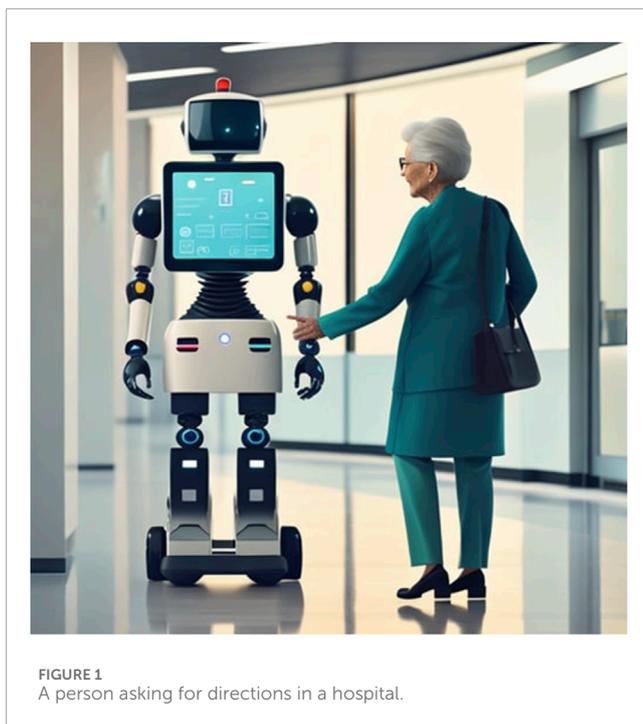


FIGURE 1  
A person asking for directions in a hospital.

The field of robotics has also seen incredible advancements over the past years, today's robots can navigate obstacle courses (Xiao et al., 2022), manipulate objects in their environment (Chai et al., 2022), generate human-like gestures (Tatarian et al., 2022), and follow complex human instructions using large language models (LLMs) (Ahn et al., 2022). Sadly, while collaboration between these two fields is common, they often begin after the robot has been designed. In our experience from multiple international robot dialogue projects, spoken interaction is not considered during the initial design phase. This lack of early-stage collaboration between robot designers and SDS researchers leaves room for oversight of critical features for spoken interaction, contributing to both performance and social errors (Tian and Oviatt, 2021). In this paper, we have combed the literature and drawn from our own experiences to highlight underlying and fundamental hardware problems that repeatedly surface when experimenting with social robots and real users. We hope this paper sparks discussion between both communities to create fully functioning social robots that do genuinely work in public spaces in the future, enabling live in-the-wild experiments.

## 2 People struggle to hear robots

The first issue that crops up commonly in the literature is the limited volume of the robot's voice. Robot designers simply attach a speaker to the robot without considering the fact that the world is noisy, and some users, such as older adults, may have hearing loss.

In recent work, researchers deployed a robot to interact with real users in an assisted living facility. The robot had to be fitted with an additional speaker that had a louder maximum volume. The modification was necessary because the users simply could not hear the robot's voice, preventing any basic interaction (Stegner et al., 2023).

This issue is not constrained to this particular setting, or to one particular robot. For example, researchers had to repeat every sentence the robot said in the lobby of a concert hall, as participants could not hear it (Langedijk et al., 2020). In various school environments, the robot's volume was not loud enough to enable effective interaction, so external speakers had to be fitted (Nikolopoulos et al., 2011). When guiding people in an elder care facility, the robot's single speaker faced the wrong direction, so users could not hear it (Langedijk et al., 2020). Another robot was deployed in the homes of a few older adults, and they noted that its volume was not loud enough. People could not hear a social robot in a gym (Sackl et al., 2022), and the list goes on.

Robots are expensive, but the speakers that researchers had to retrofit to the robots were inexpensive and readily available. This low-cost change was simple, yet *crucial*, to enable effective communication with a user in a real-world setting. When designing robots for spoken interaction, we recommend fitting multiple speakers (facing various directions) that have a loud maximum volume. This will guarantee that the robot can be heard in public spaces, and ensure its accessibility for people with limited hearing. In the future, parametric array loudspeakers (PALs) could be installed to use ultrasonic transducers (Yang et al., 2005; Bhorge et al., 2023). PALs are unidirectional, using the nonlinear interactions between soundwaves to enable directed personal communication to a specific user in a populated environment (Zhu et al., 2023).

## 3 Robots struggle to hear people

There is another conversation participant that cannot properly hear what their interlocutor is saying—the robot. This problem is similar to the one in Section 2, and is also frequently found in the literature. A social robot struggled to hear users in a hotel lobby, for example (Hahkio, 2020). Many researchers retrofit better microphones to the robot (Villalpando et al., 2018), or next to the robot (Wagner et al., 2023), in order to hear the user more clearly.

In an assisted living facility, researchers had to resort to listening to the user through an ajar door to run their experiments. The microphone array could not reliably pick up what users said (Stegner et al., 2023).

Home voice assistants do successfully hear people in noisy environments, like family homes (Porcheron et al., 2018), however. They can pick up what the user said when other conversations are happening in the room, and when the TV or radio are on [we are also finding this in ongoing work (Addlesee, 2022)]. Today's social robots typically have four microphones<sup>1</sup>, but we argue that this is far too few. Apple's Homepod originally had six microphones (Calore, 2019), and Amazon's Alexa Echo had seven (Spekking, 2021). The newest Homepod and Echo have reduced to four microphones for two reasons: (1) These devices are incentivised to keep their device's costs low to encourage adoption by new users (Welch, 2023); and (2) The device's shape and internal component arrangements have

<sup>1</sup> We checked the technical specifications of several commonly used social robots, and robots that we have deployed ourselves. We are refraining from naming specific robot creators, as this paper aims to encourage collaboration, and not criticise specific robots.

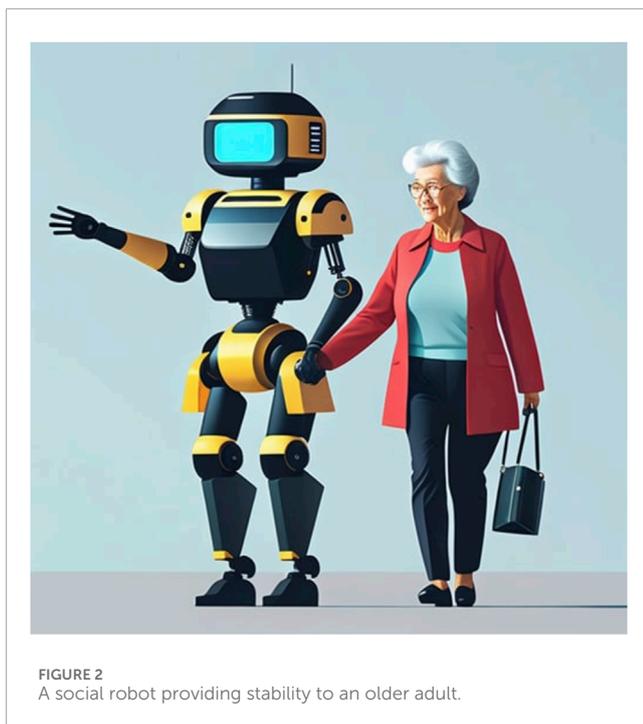


FIGURE 2  
A social robot providing stability to an older adult.

been refined and optimised over many years through experiments with millions of users (Wilson, 2020). Robots do not share either of these features. Microphones are trivially inexpensive relative to the price of a robot, and instead of helping microphones, the robot's shape actively hinders their performance. The body parts of a social robot often sit between the user and the microphone (for example, when the user is behind the robot, or in a wheelchair). Robots also create a lot of noise themselves, called *ego-noise*. Related research required high-quality audio input from a noisy propellered UAV, so they attached sixteen microphones in various locations around the device (Nakadai et al., 2017), not just four.

Human-robot spoken communication can also be disrupted by societal or linguistic phenomena, such as overlapping or poorly formed turn-taking conditions (Skantze, 2021). Such conditions include barging-in (Wagner et al., 2021) (where the user interrupts the robot mid-sentence, but the robot fails to recognise that the user started speaking), and poor end-of-turn detection (due to long pauses or intermittent speech from the user). In our experience, users sometimes barge-in because of high latency caused by limited computational power onboard the robot, or on-site connectivity issues, in addition to the SDS latency.

Potential approaches to this challenge include incremental dialogue processing (Addlesee et al., 2020; Aylett et al., 2023), predictive turn-taking (Inoue et al., 2024), or explicit turn-taking signals which enable the user to better understand when the robot is actually listening to them. For instance, Foster et al. (2019) employed a tablet on the robot's torso that was showing "I am listening" and "I am speaking" text to help guide the users in a noisy shopping mall setup.

We therefore recommend fitting multiple high-quality microphones in various locations around the robot's body, as well as using appropriate signal processing techniques, such as beamforming (Adel et al., 2012). Latency issues must be addressed within

the SDS, and by increasing the robot's computational capabilities. These changes will again ensure that multi-party spoken interactions can realistically take place in public spaces. Robot designers must also consider microphone placement lower down on the robot for shorter users, and for people in wheelchairs, as they are commonly just placed on the top of the robot's head.

### 3.1 Multi-party interaction

All of the above challenges assume that the interaction is dyadic—that is, one person conversing with a single system/robot. Conversational AI systems and SDSs are typically designed for this setting, including commercial assistants like Alexa and Siri. However, dyadic interactions can only be guaranteed in specific environments, such as single-occupant homes (and even then, there may be visitors). In the public spaces that social robots are expected to roam in the future (see Section 1), groups of people may approach the robot (Alameda-Pineda et al., 2024). In multi-party conversations (MPCs), the SDS must track who said an utterance, who the user was addressing, and then generate a suitable response, depending on whether the robot is addressing an individual or the whole group (Traum, 2004). The robot may also need to decide to remain silent, for example if people are talking to each other, but still monitor the content of their conversation in case it can assist them. Additionally, MPCs introduce unique challenges such as multi-party goal-tracking (Addlesee et al., 2023b). Groups may have conflicting goals, or share goals (Eshghi and Healey, 2016). Current social robots are not designed to enable MPCs, since speaker diarization (tracking 'who said what') is critical (Addlesee et al., 2023c; Schauer et al., 2023). The audio from the robot's microphones must not only be clear enough to perform ASR accurately, but clear enough to determine *who* said an utterance (Cooper et al., 2023). Ideally, the microphones would also provide the angle which the audio originated from. This angle can be combined with the robot's vision to determine which person in view said an utterance. The robot can then look at the user it is addressing when responding.

We recommend that social robots be designed with multi-party interaction in mind—this means designing microphone arrays such that speaker diarization is accurate, combining this with person-tracking, and developing NLP systems that can understand and manage multi-party conversations (Lemon, 2022).

## 4 Ego-noise

This issue of ego-noise, introduced in Section 3, is so problematic that an entire field of research has grown to tackle it. Researchers find that ego-noise, noise generated by the robot itself, does not just negatively impact ASR performance, but that ego-noise reduction methods also suppress some of the user's utterance (Ince et al., 2010; Schmidt et al., 2018). To clarify, both the ego-noise reduction techniques, and the ego-noise itself negatively impact ASR performance (Alameda-Pineda et al., 2024).

This issue would be helped by additional speakers and microphones, ideally not placed next to noise sources, allowing both parties to hear each other. An optimal social robot designed for spoken interaction would also have much quieter joint motors and

fans. These are more expensive than speakers and microphones, but they would greatly improve the SDSs ability to understand the user. This could be paired with research to repair and understand disrupted sentences (Addlesee and Damonte, 2023a; Addlesee and Damonte, 2023b), while quieter motors are developed.

In addition to joint motors and fans, the robot's own voice is another source of ego-noise. Microphones cannot simply be turned off when the robot is talking, as speech can be overlapping, so recognised speech may have to be classified as being produced by itself or another (Lemon and Gruenstein, 2004).

## 5 Joint robustness

Ego noise obviously does not impact robots that do not have a body. In our view, though, social robots should be able to point to location and objects, guide users, and help users physically. For example, consider a hospital waiting room in a hospital memory clinic (Gunson et al., 2022). Patients are typically older adults, and may use the robot's arm for stability, like they would with another human (see Figure 2). Current social robots can generate social gestures like waving or holding its hand out for a handshake. If you were to shake the robot's hand, however, it would likely break. Such fragility could potentially harm users if deployed in this setting. People may assume that they can link arms with the robot while being guided, a perfectly natural assumption. When an older adult puts their weight on the robot's joint, though, they might fall. This is clearly a potentially harmful design flaw that must be resolved if we are ever going to find robot assistants in the wild.

## 6 Conclusion

Interacting naturally with social robots in public spaces is currently still a sci-fi fantasy. There are challenges that SDS researchers must tackle to reach this goal, but that is not the only bottleneck. Even a perfect SDS would fail if it was embedded within today's social robots. We have highlighted that robots need louder speakers (or parametric array loudspeakers in the future), a greater number of high-quality microphones, quieter fans, and quieter motors to allow both parties to hear each other. These are critical problems that completely block spoken interactions outside a lab setting. We highlighted that robots also need to be more physically robust if they are to be safely applied in the real world, particularly in settings with older adults.

Social robotics research will continue to rely on offline evaluations, wizard-of-oz deployments, or lab-based experiments if these robot hardware issues are not resolved. Our suggestions are not an exhaustive list, but we hope that they spark discussion and encourage collaboration between robot designers and SDS researchers. This collaboration should take place in the initial stages of a robot's design to avoid

the retrofitting of hardware and sensors discussed in this paper, and instead enable real in-the-wild experiments.

## 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

AA: Conceptualization, Investigation, Writing–original draft, Writing–review and editing. IP: Writing–original draft, Writing–review and editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was partially supported by the European Commission under the Horizon 2020 framework programme for Research and Innovation (H2020-ICT-2019-2, GA no. 871245): SPRING project <https://spring-h2020.eu/>.

## Acknowledgments

The images in this paper were generated with Hotpot.io.

## Conflict of interest

Author IP was employed by company Aveni AI. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor FF and reviewer DK declared a past co-authorship with the author IP.

## 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

Addlesee, A. (2022). "Securely cpeople's interactions with voice assistants at home: a bespoke tool for ethical data collection," in *Proceedings of the second workshop on NLP for positive impact (NLP4PI)*, 25–30.

Addlesee, A. (2023). "Voice assistant accessibility," in *The international workshop on spoken dialogue systems Technology, IWSDS 2023*, 11.

- Addlesee, A., and Damonte, M. (2023a). *Understanding disrupted sentences using underspecified abstract meaning representation*. Dublin, Ireland: Interspeech, 5.
- Addlesee, A., and Damonte, M. (2023b). Understanding and answering incomplete questions in Proceedings of the 5th conference on conversational user interfaces, 9, 1–9. doi:10.1145/3571884.3597133
- Addlesee, A., Denley, D., Edmondson, A., Gunson, N., Garcia, D. H., Kha, A., et al. (2023c). "Detecting agreement in multi-party dialogue: evaluating speaker diarisation versus a procedural baseline to enhance user engagement," in *Proceedings of the workshop on advancing GROUp UNderstanding and robots aDaptive behaviour (GROUND)*, 7.
- Addlesee, A., Sieińska, W., Gunson, N., Garcia, D. H., Dondrup, C., and Lemon, O. (2023b). Multi-party goal tracking with LLMs: comparing pre-training, fine-tuning, and prompt engineering in *Proceedings of the 24th annual meeting of the special interest group on discourse and dialogue*, 13.
- Addlesee, A., Sieińska, W., Gunson, N., Hernández García, D., Dondrup, C., and Lemon, O. (2023a). Data collection for multi-party task-based dialogue in social robotics. *Int. Workshop Spok. Dialogue Syst. Technol. IWSDS 2023*, 10. doi:10.21437/interspeech.2023-307
- Addlesee, A., Yu, Y., and Eshghi, A. (2020). A comprehensive evaluation of incremental speech recognition and diarization for conversational AI in Proceedings of the 28th international conference on computational linguistics, 3492–3503.
- Adel, H., Souad, M., Alaqeeli, A., and Hamid, A. (2012). Beamforming techniques for multichannel audio signal separation 6, 659, 667. doi:10.4156/jdcta.vol6.issue20.72
- Ahn, M., Brohan, A., Brown, N., Chebotar, Y., Cortes, O., David, B., et al. (2022). Do as i can, not as i say: grounding language in robotic affordances. arXiv Prepr. arXiv:2204.01691. doi:10.48550/arXiv.2204.01691
- Alameda-Pineda, X., Addlesee, A., García, D. H., Reinke, C., Arias, S., Arrigoni, F., et al. (2024). Socially pertinent robots in gerontological healthcare. arXiv preprint arXiv:2404.07560.
- Aylett, M. P., Carmantini, A., and Braude, D. A. (2023). Why is my social robot so slow? How a conversational listener can revolutionize turn-taking in Proceedings of the 5th conference on conversational user interfaces, 4.
- Bhorge, S., Patil, V., Pawar, S., Poke, A., and Jambhulkar, R. (2023). Unidirectional parametric speaker. ITM Web Conf. ITM Web of Conferences Les Ulis, France: EDP Sciences, 56, 04010, doi:10.1051/itmconf/20235604010
- Calore, M. (2019). Review: apple Homepod. *Wired*.
- Chai, H., Li, Y., Song, R., Zhang, G., Zhang, Q., Liu, S., et al. (2022). A survey of the development of quadruped robots: joint configuration, dynamic locomotion control method and mobile manipulation approach. *Biomim. Intell. Robotics 2*, 100029. doi:10.1016/j.birob.2021.100029
- Cooper, S., Ros, R., Lemaignan, S., and Robotics, P. (2023). Challenges of deploying assistive robots in real-life scenarios: an industrial perspective in The 32nd IEEE international conference on robot and human interactive communication. Naples, Italy: RO-MAN, 10.
- Eshghi, A., and Healey, P. G. (2016). Collective contexts in conversation: grounding by proxy. *Cognitive Sci.* 40, 299–324. doi:10.1111/cogs.12225
- Foster, M. E., Craenen, B., Deshmukh, A., Lemon, O., Bastianelli, E., Dondrup, C., et al. (2019). *MUMMER: socially intelligent human-robot interaction in public spaces*.
- Glass, J. (1999). Challenges for spoken dialogue systems. *Proc. 1999 IEEE ASRU Workshop (MIT Laboratory Comput. Sci. Camb. MA, USA)* 696, 10.
- Gunson, N., Garcia, D. H., Sieińska, W., Addlesee, A., Dondrup, C., Lemon, O., et al. (2022). "A visually-aware conversational robot receptionist," in *Proceedings of the 23rd annual meeting of the special interest group on discourse and dialogue*, 645–648.
- Hahkio, L. (2020). *Service robots' feasibility in the hotel industry: a case study of Hotel Presidentti*. Vantaa, Finland: Laurea University of Applied Sciences.
- Honig, S., and Oron-Gilad, T. (2018). Understanding and resolving failures in human-robot interaction: literature review and model development. *Front. Psychol.* 9, 861. doi:10.3389/fpsyg.2018.00861
- Ince, G., Nakadai, K., Rodemann, T., Hasegawa, Y., Tsujino, H., and Ji, I. (2010). A hybrid framework for ego noise cancellation of a robot in 2010 IEEE international Conference on Robotics and automation. IEEE, 3623–3628.
- Inoue, K., Jiang, B., Ekstedt, E., Kawahara, T., and Skantze, G. (2024). Multilingual turn-taking prediction using voice activity projection in Proceedings of the 2024 joint international conference on computational linguistics, language resources and evaluation (LREC-COLING 2024), 11873–11883. Available at: <https://aclanthology.org/2024.lrec-main.1036/>
- Langedijk, R. M., Odabasi, C., Fischer, K., and Graf, B. (2020). Studying drink-serving service robots in the real world in 2020 29th IEEE international conference on robot and human interactive communication. IEEE, 788–793.
- Lemon, O. (2022). Conversational AI for multi-agent communication in natural language. *AI Commun.* 35, 295–308. doi:10.3233/aic-220147
- Lemon, O., and Gruenstein, A. (2004). Multithreaded context for robust conversational interfaces: context-sensitive speech recognition and interpretation of corrective fragments. *ACM Trans. Computer-Human Interact. (TOCHI)* 11, 241–267. doi:10.1145/1017494.1017496
- Nakadai, K., Kumon, M., Okuno, H. G., Hoshiba, K., Wakabayashi, M., Washizaki, K., et al. (2011). Development of microphone-array-embedded UAV for search and rescue task in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) IEEE, 5985–5990.
- Nikolopoulos, C., Kuester, D., Sheehan, M., Ramteke, S., Karmarkar, A., Thota, S., et al. (2011). *Robotic agents used to help teach social skills to children with autism: the third generation*. IEEE, 253–258.
- Porcheron, M., Fischer, J. E., Reeves, S., and Sharples, S. (2018). "Voice interfaces in everyday life," in Proceedings of the 2018 CHI conference on human factors in computing systems, 1–12.
- Sackl, A., Pretolesi, D., Burger, S., Ganglbauer, M., and Tscheligi, M. (2022). Social robots as coaches: how human-robot interaction positively impacts motivation in sports training sessions in 2022 31st IEEE international conference on robot and human interactive communication. IEEE, 141–148.
- Schauer, L., Sweeny, J., Lyttle, C., Said, Z., Szeles, A., Clark, C., et al. (2023). Detecting agreement in multi-party conversational AI in *Proceedings of the workshop on advancing GROUp UNderstanding and robots aDaptive behaviour (GROUND)*, 5.
- Schmidt, A., Löllmann, H. W., and Kellermann, W. (2018). A novel ego-noise suppression algorithm for acoustic signal enhancement in autonomous systems in 2018 IEEE international Conference on acoustics, Speech and signal processing (ICASSP) IEEE, 6583–6587.
- Skantze, G. (2021). Turn-taking in conversational systems and human-robot interaction: a review. *Comput. Speech and Lang.* 67, 101178. doi:10.1016/j.csl.2020.101178
- Spekking, R. (2021). *Amazon Echo dot (RS03QR) - LED and microphone board*. Amsterdam, Netherlands: Wikimedia.
- Stegner, L., Senft, E., and Mutlu, B. (2023). Situated participatory design: a method for *in situ* design of robotic interaction with older adults in Proceedings of the 2023 CHI conference on human factors in computing systems, 1–15.
- Tatarian, K., Stower, R., Rudaz, D., Chamoux, M., Kappas, A., and Chetouani, M. (2022). How does modality matter? investigating the synthesis and effects of multi-modal robot behavior on social intelligence. *Int. J. Soc. Robotics* 14, 893–911. doi:10.1007/s12369-021-00839-w
- Thrun, S. (1998). When robots meet people. *IEEE Intelligent Syst. their Appl.* 13, 27–29. doi:10.1109/5254.683178
- Tian, L., and Oviatt, S. (2021). A taxonomy of social errors in human-robot interaction. *ACM Trans. Human-Robot Interact. (THRI)* 10, 1–32. doi:10.1145/3439720
- Tolmeijer, S., Weiss, A., Hanheide, M., Lindner, F., Powers, T. M., Dixon, C., et al. (2020). Taxonomy of trust-relevant failures and mitigation strategies. *Proc. 2020 acm/ieee Int. Conf. human-robot Interact.*, 3–12. doi:10.1145/3319502.3374793
- Traum, D. (2004). Issues in multiparty dialogues. Advances in agent communication: international workshop on agent communication languages, ACL 2003 in *Revised and invited papers*. Melbourne, Australia: Springer, 201–211.
- Villalpando, A. P., Schillaci, G., and Hafner, V. V. (2018). Predictive models for robot ego-noise learning and imitation in 2018 joint IEEE 8th international Conference on Development and Learning and epigenetic robotics (ICDL-EpiRob) IEEE, 263–268.
- Wagner, N., Kraus, M., Lindemann, N., and Minker, W. (2023). Comparing multi-user interaction strategies in human-robot teamwork. *Int. Workshop Spok. Dialogue Syst. Technol. IWSDS 2023*, 12.
- Wagner, N., Kraus, M., Rach, N., and Minker, W. (2021). *How to address humans: system barge-in in multi-user HRI*. Springer Singapore, 147–152. doi:10.1007/978-981-15-9323-9\_13
- Welch, C. (2023). Apple's new HomePod unsurprisingly sounds close to the original. *Verge*.
- Williams, J. D. (2009). *Spoken dialogue systems: challenges, and opportunities for research*. Merano, Italy: ASRU, 25.
- Wilson, M. (2020). *Why Amazon radically redesigned the Echo*. New York, NY: Fast Company.
- Xiao, X., Xu, Z., Wang, Z., Song, Y., Warnell, G., Stone, P., et al. (2022). Autonomous ground navigation in highly constrained spaces: lessons learned from the benchmark autonomous robot navigation challenge at ICRA 2022. *IEEE Robotics and Automation Mag.* 29, 148–156. doi:10.1109/mra.2022.3213466
- Yang, J., Gan, W. S., Tan, K. S., and Er, M. H. (2005). Acoustic beamforming of a parametric speaker comprising ultrasonic transducers. *Sensors Actuators A Phys.* 125, 91–99. doi:10.1016/j.sna.2005.04.037
- Zhu, Y., Ma, W., Kuang, Z., Wu, M., and Yang, J. (2023). Optimal audio beam pattern synthesis for an enhanced parametric array loudspeaker. *J. Acoust. Soc. Am.* 154, 3210–3222. doi:10.1121/10.0022415