Introduction to the Special Issue on Sound in Human-Robot Interaction

FREDERIC ROBINSON, University of New South Wales, Australia
HANNAH PELIKAN, Linköping University, Sweden
KATSUMI WATANABE, Waseda University, Japan
LUISA DAMIANO, IULM University, Italy
OLIVER BOWN and MARI VELONAKI, University of New South Wales, Australia

ACM Reference format:
Frederic Robinson, Hannah Pelikan, Katsumi Watanabe, Luisa Damiano, Oliver Bown, and Mari Velonaki. 2023. Introduction to the Special Issue on Sound in Human-Robot Interaction. *ACM Trans. Hum.-Robot Interact.* 12, 4, Article 45 (December 2023), 5 pages.
https://doi.org/10.1145/3632185

1 INTRODUCTION

Sound is an important interaction modality and a large part of human interaction happens in the aural domain. While research in **Human-Robot Interaction** (**HRI**) has long explored spoken language for interacting with humans, sound as a broader—and, to a significant degree, non-lexical (i.e., without words)—medium has been given comparably less attention. Yet, the range of sounds that robots can produce is vast, encompassing, among others, mechanical noise, music, and utterances that mimic human and animal vocalizations with varying degrees of realism. The sound of a robot's machinery can shape our perceptions and expectations [11, 17], music serves as a medium for robots to engage and communicate [18], and shared musical experiences can strengthen the bond between humans and robots [5]. Sonifications may enhance the legibility of movement and gestures [4, 7, 15] and beep sounds may be used to communicate emotions [2, 14]. Getting closer to the margins of language, robots may take inspiration from non-lexical fillers such as "uh" [13, 16] and backchannels such as "mhmm" [8, 12]. More generally, pitch, intensity, and other human prosodic variations may be drawn on in robot sound design [3, 10]. The information that can be extracted from sound in a robot's environment is equally rich. Beyond the recognition of semantic content, robots use, for example, sound source localization to gain a better understanding of their environment [9], or analyze a human's voice timbre and tone to distinguish speakers [6] and detect emotion [1].

Authors' addresses: F. Robinson, O. Bown, and M. Velonaki, University of New South Wales, Cnr Oxford St & Greens Rd, Paddington NSW 2021, Australia; e-mails: frederic.robinson@unsw.edu.au, o.bown@unsw.edu.au, mari.velonaki@unsw.edu.au; H. Pelikan, Linköping University, SE-581 83 Linköping, Sweden; e-mail: hannah.pelikan@liu.se; K. Watanabe, Waseda University, 1-chōme-104 Totsukamachi, Shinjuku City, Tokyo 169-8050, Japan; e-mail: katz@waseda.jp; L. Damiano, IULM University, Via Carlo Bo, 1, 20143 Milano MI, Italy; e-mail: luisa.damiano@iulm.it.

2573-9522/2023/12-ART45
https://doi.org/10.1145/3632185

The disciplines involved in these research pursuits are diverse, ranging from engineering, music, sound, and sonic interaction design to psychology, linguistics, and conversation analysis. However, there has so far been little knowledge exchange between the different disciplines despite them often tackling similar challenges around robot sound. This special issue aims to take a step towards unifying the Sound in HRI community, celebrating and showcasing the diversity of the different approaches while highlighting points of convergence.

While there are existing works that take a broader perspective on certain sound-related subfields, such as semantic-free utterances [19] or robotic musicianship, a central resource that provides a comprehensive overview of these diverse disciplines, and how they interact with each other, has yet to be created. This Special Issue is, to our knowledge, the first to address this by providing researchers with a broad picture of how the medium of sound can be utilized to enrich and refine interactions between robots and humans. The issue is meant for social robotics researchers and practitioners across academia and industry, both as an introduction for readers interested in entering the field, as well as a summary of findings and best practices for readers with prior experience. By showcasing the various ways in which sound informs, influences, and engages people, we aim to provide readers with new ways of using this modality to create richer, more nuanced HRIs.

2 OBJECTIVE

When considering the use of sound in HRI, it is important to note that (i) some form of sound is always present during interactions, i.e., due to their material presence robots are not silent, and (ii) listeners continuously interpret and make sense of sound in some way. Sound can provide important insights both for robots and their operators: Humans draw on auditory cues when making assumptions about a robot's characteristics, intentions, and capabilities. These cues occur not in isolation but are also tightly coupled to a robot's body and movement, both of which are equally perceptually relevant. Designers and researchers should, therefore, carefully consider the soundscapes accompanying the HRIs they are designing and ensure that any information conveyed through this channel is transmitted deliberately.

In light of these considerations, this Special Issue aims to support HRI researchers adopting a holistic perspective on sound in HRI, and making thoughtful and deliberate choices on how the medium of sound is used in their work. We argue that a well-considered utilization of sound will result in more successful designs than an ill-considered one that does not take into account the wide array of challenges and opportunities that this modality offers. The Special Issue pursues three objectives: First, to provide the reader with a comprehensive overview of how sound is, and could be applied in HRI research. Second, to present a collection of studies exploring the medium of sound in a multitude of application contexts, from the sonification of humanoid robots and drones, across the timing and timbre of utterances, to the extraction of relevant information from background noise. Third, to showcase a cross-section of the field's interdisciplinary research pursuits, providing a collective resource of research directions, methodologies, and use cases that will inspire future work in the field.

3 THE ARTICLES

This Special Issue contains 11 articles investigating sound in HRI. Two of them provide the reader with an overview of the field and present comprehensive conceptual frameworks that can be used to gain a holistic perspective of the modality. Another two articles showcase how a robot can extract relevant information from its sonic environment through the detection of social presence and the use of spatial audio in teleoperation contexts. Four articles present robot sonification strategies for drones, robot swarms, robotic arms, and humanoid robots, with goals including the conveyance of emotion and the creation of a general sonic presence. The final three articles refine robot

communication by embedding additional information within it, specifically exploring the effects of timing and timbre of utterances. The articles can be summarized as follows.

Brian J. Zhang and Naomi T. Fitter's "Nonverbal Sound in HRI: a Systematic Review" presents an overview of the field through a systematic review of 148 peer-reviewed articles to identify and compare different use cases and approaches for incorporating nonverbal sound in HRI. This work results in two taxonomies and recommendations for the design, generation, and validation of nonverbal robot sound.

Focusing on sound perception, two articles look at refining a robot's and robot operator's awareness of its sonic environment. Nicholas Georgiou and colleagues in their article "Is Someone There Or Is That The TV? Detecting Social Presence Using Sound" investigates the use of sound-based classification to improve interaction capabilities of social robots in a home environment, distinguishing between natural conversation from users and speech content from media playback. Utilizing a unique dataset of different acoustic environments, the authors assessed various machine learning classifiers and found the C-**Support Vector Classification** (**SVC**) algorithm to be most effective, ultimately proposing a sound classification pipeline for home robots to better engage with users. Adam K. Coyne and colleagues in their article ""Who said that?" Applying the Situation Awareness Global Assessment Technique to Social Telepresence" introduces an objective measurement, based on the **Situation Awareness Global Assessment Technique** (**SAGAT**), to evaluate operator situation awareness during the teleoperation of social robots. They apply this technique to evaluate the impact of different audio feedback on operator situation awareness. While the initial study showed no significant differences between mono- and binaural audio feedback, correlations among measures were noted, indicating the potential for developing specialized assessment techniques for social situation awareness in teleoperated robots and guiding future robot design decisions.

Focusing on sound generation, Bastian Orthmann and colleagues' "Sounding Robots: Design and Evaluation of Auditory Displays for Unintentional HRI" works towards a holistic framework of implicit robot communication by presenting a multi-layer sound-based classification system designed to communicate various states and intentions of robots in a shared space with humans, such as urgency, availability, and directionality. Through a series of online studies, they found that the created sounds were generally recognized as intended by participants, suggesting that sounds can be an effective tool for intuitive and implicit communication of robot states and intended actions.

Four articles present sonification strategies for a range of robotic agents. Ziming Wang and colleagues' "The Effects of Natural Sounds and Proxemic Distances on the Perception of a Noisy Domestic Flying Robot" investigates the impact of adding natural sounds, specifically birdsong and rain, to the consequential noise of domestic drones during close-range interactions with humans, with a mixed-methods study showing that these sounds and proxemic distances significantly influence user perceptions. Moreover, findings show that user perceptions are also strongly influenced by their past experiences, leading to six concrete design recommendations for sound implementation in domestic drones. Elias Naphausen and colleagues' "New Design Potentials of Non-mimetic Sonification in HRI" details a project exploring the potential of non-mimetic sonification (using sound properties like pitch, volume, and timbre) to improve HRI, leveraging data from a 7-axis manipulator to create an augmented audible presence and enable new forms of interaction. It presents research parameters, an empirical study setup, and potential implications of integrating these sonification findings into a unified HRI process, particularly focusing on the interplay between machinic and auditory dimensionality. Adrian B. Latupeirissa and colleagues' "Probing Aesthetics Strategies for Robot Sound: Complexity and Materiality in Movement Sonification" explores the aesthetic impact of the sonification of a Pepper robot's movements across three studies

using two sets of sound models. Findings suggest that participants preferred more complex sound models and subtle sounds that blend well with ambient noise, with sound preferences influenced by the context in which the robot-generated sounds were experienced. Maria Mannone and colleagues' "The Sound of Swarm. Auditory Description of Swarm Robotic Movements" presents a theoretical framework linking musical parameters (pitch, timbre, loudness, and articulation) with robotic parameters (position, identity, motor status, and sensor status) to facilitate communication within a robotic swarm through sound. Utilizing Hilbert spaces, the framework enables quantum representations of musical states, presenting potential applications through case studies involving simulated scenarios with robo-caterpillars, robo-ants, and robo-fish.

Finally, two articles take inspiration from how humans use vocal sounds and explore the design of robot utterances. Xiaozhen Liu and colleagues' "Robots" "Woohoo" and "Argh" Can Enhance Users' Emotional and Social Perceptions: An Exploratory Study on Non-Lexical Vocalizations and Non-Linguistic Sounds' investigates how sounds can convey basic emotions in a humanoid robot, Pepper, and how this influences user perception. The article explores the interplay between different vocalizations, non-linguistic utterances, and robot gestures. Findings indicate that vocalizations produced with a natural voice resulted in significantly higher emotion recognition accuracy and induced higher trust, naturalness, and preference, while musical sounds showed lower perception ratings. Kerstin Fischer and Oliver Niebuhr's "Which Voice for which Robot? Designing Robot Voices that Indicate Robot Size" investigates the correlation between human voice acoustic parameters and body size, with an aim to design suitable voices for robots of varying sizes. It discovers certain acoustic features significantly linked with body height and weight, which when applied to robotic speech, can reliably cue a listener to the perceived size of the robot.

We are delighted to highlight the breadth and diversity of disciplines, methodologies, and application contexts represented in this Special Issue on Sound in HRI. The included articles reflect the rich tapestry of this multidisciplinary field and showcase examples of the work currently being done on robot sound perception and production for different platforms and contexts. We would like to express our sincere appreciation to the reviewers who have dedicated their time and expertise to ensuring the quality of this publication. Their rigorous evaluation and insightful feedback have significantly contributed to this Special Issue. It is our hope that this collection of work offers an insightful and inspiring overview of this exciting area of study that can serve both as an introduction to the topic of Sound in HRI and as a resource for identifying new research directions and fostering interdisciplinary collaborations.

REFERENCES

[1] Fernando Alonso-Martin, Maria Malfaz, Joao Sequeira, Javier F. Gorostiza, and Miguel A. Salichs. 2013. A multimodal emotion detection system during human-robot interaction. *Sensors* 13, 11 (2013), 15549–15581.

[2] Lilian Chan, Brian J Zhang, and Naomi T Fitter. 2021. Designing and validating expressive cozmo behaviors for accurately conveying emotions. In *Proceedings of the 2021 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1037–1044.

[3] Kerstin Fischer, Lars Christian Jensen, and Leon Bodenhagen. 2014. To beep or not to beep is not the whole question. In *Proceedings of the Social Robotics: 6th International Conference, ICSR 2014*. 156–165. DOI: https://doi.org/10.1007/978-3-319-11973-1_16

[4] Emma Frid and Roberto Bresin. 2022. Perceptual evaluation of blended sonification of mechanical robot sounds produced by emotionally expressive gestures: Augmenting consequential sounds to improve non-verbal robot communication. *International Journal of Social Robotics* 14, 2 (2022), 357–372. DOI: https://doi.org/10.1007/s12369-021-00788-4

[5] Guy Hoffman, Shira Bauman, and Keinan Vanunu. 2016. Robotic experience companionship in music listening and video watching. *Personal and Ubiquitous Computing* 20, 1 (2016), 51–63. DOI: https://doi.org/10.1007/s00779-015-0897-1

[6] Mikyong Ji, Sungtak Kim, Hoirin Kim, and Ho-Sub Yoon. 2008. Text-independent speaker identification using soft channel selection in home robot environments. *IEEE Transactions on Consumer Electronics* 54, 1 (2008), 140–144. DOI: https://doi.org/10.1109/TCE.2008.4470036

[7] Gunnar Johannsen. 2001. Auditory displays in human-machine interfaces of mobile robots for non-speech communication with humans. *Journal of Intelligent and Robotic Systems* 32, 2 (2001), 161–169. DOI: https://doi.org/10.1023/A:1013953213049

[8] Malte F. Jung, Jin Joo Lee, Nick DePalma, Sigurdur O. Adalgeirsson, Pamela J. Hinds, and Cynthia Breazeal. 2013. Engaging robots: Easing complex human-robot teamwork using backchanneling. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. 1555–1566.

[9] Xiaofei Li and Hong Liu. 2012. Sound source localization for HRI using FOC-based time difference feature and spatial grid matching. *IEEE Transactions on Cybernetics* 43, 4 (2012), 1199–1212.

[10] Angelica Lim and Hiroshi G. Okuno. 2014. The MEI robot: Towards using motherese to develop multimodal emotional intelligence. *IEEE Transactions on Autonomous Mental Development* 6, 2 (2014), 126–138. DOI: https://doi.org/10.1109/TAMD.2014.2317513

[11] Dylan Moore, Hamish Tennent, Nikolas Martelaro, and Wendy Ju. 2017. Making noise intentional: A study of servo sound perception. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*. ACM Press, Vienna, Austria, 12–21. DOI: https://doi.org/10.1145/2909824.3020238

[12] Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. 2017. Telling stories to robots: The effect of backchanneling on a child's storytelling. In *Proceedings of the 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 100–108. DOI: https://doi.org/10.1145/2909824.3020245

[13] Hannah Pelikan and Emily Hofstetter. 2023. Managing delays in human-robot interaction. *ACM Transactions on Computer-Human Interaction* 30, 4, Article 50 (2023), 42 pages. DOI: https://doi.org/10.1145/3569890

[14] Robin Read and Tony Belpaeme. 2012. How to use non-linguistic utterances to convey emotion in child-robot interaction. In *Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '12*. ACM, 219. DOI: https://doi.org/10.1145/2157689.2157764

[15] Richard Savery, Ryan Rose, and Gil Weinberg. 2019. Establishing human-robot trust through music-driven robotic emotion prosody and gesture. In *Proceedings of the 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1–7.

[16] Toshiyuki Shiwa, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, and Norihiro Hagita. 2009. How quickly should a communication robot respond? Delaying strategies and habituation effects. *International Journal of Social Robotics* 1, 2 (2009), 141–155.

[17] Hamish Tennent, Dylan Moore, Malte Jung, and Wendy Ju. 2017. Good vibrations: How consequential sounds affect perception of robotic arms. In *Proceedings of the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Lisbon, 928–935. DOI: https://doi.org/10.1109/ROMAN.2017.8172414

[18] Gil Weinberg, Mason Bretan, Guy Hoffman, and Scott Driscoll. 2020. *Robotic Musicianship: Embodied Artificial Creativity and Mechatronic Musical Expression*. Springer Nature.

[19] Selma Yilmazyildiz, Robin Read, Tony Belpeame, and Werner Verhelst. 2016. Review of semantic-free utterances in social human-robot interaction. *International Journal of Human-Computer Interaction* 32, 1 (2016), 63–85. DOI: https://doi.org/10.1080/10447318.2015.1093856