Robust Non-Verbal Expression in Humanoid Robots: New Methods for Augmenting Expressive Movements with Sound

Roberto Bresin*
roberto@kth.se
KTH Royal Institute of Technology
Sweden

Emma Frid*
emmafrid@kth.se
IRCAM STMS Lab
KTH Royal Institute of Technology
Sweden

Adrian B. Latupeirissa*
ablat@kth.se
KTH Royal Institute of Technology
Sweden

Claudio Panariello*
claudiop@kth.se
KTH Royal Institute of Technology
Sweden

ABSTRACT

The aim of the SONAO project is to establish new methods based on sonification of expressive movements for achieving a robust interaction between users and humanoid robots. We want to achieve this by combining competences of the research team members in the fields of social robotics, sound and music computing, affective computing, and body motion analysis. We want to engineer sound models for implementing effective mappings between stylized body movements and sound parameters that will enable an agent to express high-level body motion qualities through sound. These mappings are paramount for supporting feedback to and understanding robot body motion. The project will result in the development of new theories, guidelines, models, and tools for the sonic representation of high-level body motion qualities in interactive applications. This work is part of the growing research field known as data sonification, in which we combine methods and knowledge from the fields of interactive sonification, embodied cognition, multisensory perception, non-verbal and gestural communication in robots.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design**.

KEYWORDS

Sonic Interaction Design for Robots, Sonification, Sound and Music Computing, Affective Computing, Expressive Gestures

1 INTRODUCTION

In the field of social robotics, it is expected that robots could provide assistance to the elderly, people with motion impairment, and be used in educational as well as service environments. To be able to establish long-term relationships with users, it is also expected that the robots will be equipped with social interaction skills. However, expression capabilities in current humanoid robots are still limited. One of the limitations is a result of the low number of available degrees of freedoms (DoF) of robots compared to humans. In the light of this limitation, there is a need for experimenting and developing new ways and processes of human-robot interaction in order to optimize the social robot's added value. This is desired specially when considering that the way in which humans explore the world and communicate with each other is multimodal: body motion, touch, haptics, and sounds are used in combination to make sense of and react to the environment.

In this position paper we summarize the methods used in the ongoing SONAO project[1]. In this project we want to leverage multimodality of human communication and develop new methods for achieving robust and fluent interactions between users and robots [11]. The SONAO project aims to improve the comprehensibility of robot non-verbal communication (NVC) based on sonification of simplified movements to achieve a robust interaction between users and humanoid robots. The purpose is to compensate for limitations in robot communicative channels with an increased clarity of NVC through expressive movements and non-verbal sounds. To achieve this, we aim to combine knowledge from the fields of social robotics, sound and music computing, affective computing, and body motion analysis. This work is part of the growing research field known as data sonification, in which we combine methods and knowledge from the fields of interactive sonification, embodied (music) cognition, multisensory perception, non-verbal and gestural communication in robots.

2 METHODS

This section briefly summarizes the methodological framework used within the SONAO project.

2.1 Expressive Movements: From Stylized Human Movements to Robot Movements

An important aspect of the SONAO project is to use expressive movements and sonification [3, 6, 8] to compensate for the reduced DoF in which robots can express themselves. This was initially explored through the creation of a database of expressive (stylized) human movements mapped to a humanoid robot and a virtual agent (i.e. an animated robot character), in work carried out by Simon Alexanderson and colleagues [1]. The work focused on full body motion capture of a mime actor (i.e. an expert in non-verbal communication techniques) interacting in short dialogues, varying the intensity level of the non-verbal expression along five different

*All authors contributed equally to this research.

[1] https://www.kth.se/hct/mid/research/smc/projects/sonao

dimensions (e.g. frustration and joy). Findings suggested that the mapping of body movements from human movements to a virtual agent tended to preserve the recognition of emotional renderings originally performed by the mime actor. A motion re-targeting framework focusing on translating the same human movements to a NAO robot was explored in a work coordinated by Iolanda Leite [18]. The work focused on motion sequences that addresses the NAO robot's movement constraints while preserving the emotional expression of the human gesture, exploring aspects such as joint angle and angular velocities.

2.2 Perception of Sounds Inherent to Robot Movement and Blended Sonification

The next part of the SONAO project focuses on enhancing the stylized gestures described in Section 2.1 with sonifications. To do this with a non-virtual robot, e.g. a NAO, one has to also consider the sounds that are inherently produced by the robot, as this may influence human-robot interaction. In our previous work [4], we investigated the perception of sounds produced by the expressive movements developed in [1, 18], using a NAO robot. The study focused on quantitative ratings of emotions conveyed by mechanical robot sounds. Results suggested that certain mechanical sounds did convey emotional characteristics when presented in an auditory-only condition and that the sounds generally communicated arousal more effectively than valence. However, the sounds produced as a result of the NAO robot's expressive movements were not clearly coupled to the emotional reactions associated with respective gestures; for example, sounds produced by a joyful gesture appeared to convey a sensation of frustration.

In a follow-up experiment, blended sonification [17] was used to enhance and further clarify the emotional expression of sounds inherently produced by the robot movements presented in [4]. In this work, mechanical sounds of the robot were blended with sonifications. The sonification design was informed by previous research on emotional expressive communication in sound and music [2, 5]. Results from a perceptual experiment with 30 participants revealed that the blended sonifications successfully contributed to enhancement of the emotional message for sound models designed to convey frustration and joy, suggesting that blended sonification can successfully improve communication of emotions through robot sounds in auditory-only conditions.

2.3 Vocal Sketching of Stylized Movements

To provide a palette of sound models suitable for the sonification of expressive robot body movements, a study was conducted using vocal sketching as a prototyping tool for exploration of sound design in HRI. Vocal sketching involves the use of the voice and body to demonstrate the relationship between actions and sonic feedback [14] and has successfully been used in a wide range of different projects, for example in SkAT-VG [13].

Our study presented in [12] explores how notation developed for the representation of sound-based musical structures could be used for the transcription of vocal sketches representing expressive robot movements. The same mime actor from the study presented in Section 2.1 was shown muted videos of his own movements, and asked to vocalize sounds that expressed these movements. The vocal sketches were then transcribed by two composers using a sound-based notation system developed in [16]. The same composers later synthesized new sonic sketches from the annotated data, the results were then compared to investigate how the audible outcome changes for different transcriptions and synthesis routines. Another outcome was a set of sound synthesis models which can be used as a first sonic sketch for designing new sound feedback produced by robot movements.

2.4 Analysis of Robot Sound in Films

It has been suggested that human interaction with technology in science-fiction films offer lessons to interaction designers, as they reflect current understandings in terms of expectations from users [15]. Similarly, our concept of robots and how we can interact with them can be influenced by the image of robots from science fiction [7]. In order to gain understanding of sonic characteristics of robot sounds in film, we carried out a study to analyse robot sounds in films [9]. Excerpts from five films representing each decade from 1970s to 2010s were annotated and analysed based on their sonic spectra using Long Time Average Spectrum (LTAS). As an overall observation, we found that robot sonic presence is highly related to the physical appearance of robots, both to their metallic physical qualities and their body movement. Moreover, movement sounds are also used to emphasize emotions in this context (e.g. sadness is displayed by the robot by a mechanical sound characterized by a falling pitch and the character's head facing downwards). Spectral characteristics of robot voices also show significant differences compared to voices of human characters; fundamental frequency of robotic voices is either shifted to higher or lower values, and the voices span over a broader frequency band. These findings could be of relevance for future work focused on social robotics and sonification of expressive robot movement.

2.5 Perception of Different Sound Designs

In a recent study presented in [10] we explored how different sets of sounds designed for expressive robot movements of a Pepper robot can influence the perception of emotional intentions. We carried out a perceptual rating experiment with two sound sets: the first was based on sawtooths, and the second one was made up of more complex sounds based on feedback chains. The study used video recordings of a Pepper robot, in which the sounds inherently produced by the robot were preserved and mixed with new sound synthesis. Preliminary results show that the sound set based on complex sounds were generally the most preferred one and regarded as more pleasant by test participants. Results also show that sawtooth-based sounds might be suitable for representing certain emotion such as sadness and excitement. The excited movement stimulus was the less likely to be misunderstood, independently from the sound design used, probably because of the prominence of the original sound of the motors always embedded in the synthesized sounds. This suggests that for achieving a clearer communication of robot expressions characterized by a low arousal, one should make use of either masking sounds or strategically placed loudspeakers depending on the mechanical activity of the robot.

3 DISCUSSION

The combination of methods and knowledge from the fields of interactive sonification (Sections 2.2 and 2.5), embodied cognition (Sections 2.1 and 2.3), multisensory perception (Sections 2.2 and 2.5), non-verbal and gestural communication in robots (Sections 2.1, 2.2, 2.4 and 2.5) as applied so far in the SONAO project, will contribute to provide new insights for the design of sounds for HRI. We believe that the strength of our project is this combination of different fields for achieving a sonic interaction design for robots that is grounded on principles of perception and communication of expression. This will help users to better understand robot's intentions thus enhancing the overall quality of the interaction.

The sound models developed in our studies and tested with specific robots (namely NAO and Pepper Robot) have been designed using tools and methods that allow them to be scaled to other robot bodies as well. This can be done by remapping the parameters of the synthesis models. We plan to verify this with other robots, such as Miro [2].

Future work within the SONAO project will aim to further explore vocal sketching as a prototyping tool for exploration of sound design in HRI, as described in Section 2.3. Further investigations into the expectations on robot sounds through studies on science-fiction films is also being conducted.

Moreover, further use of sonification in HRI is also of our interest. Future scenarios that we will investigate through sonic interaction include the communication of robot intentions, of its real-time internal state (e.g. sensor data of the robot, internal software processes) and the display of external information (e.g. weather data, health data).

4 ACKNOWLEDGMENTS

This study was funded by grants to Roberto Bresin from KTH EECS and from the Swedish Research Council (Grant 2017-03979), and by NordForsk Nordic University Hub "Nordic Sound and Music Computing Network - NordicSMC" (project number 86892).

REFERENCES

[1] Simon Alexanderson, Carol O'sullivan, Michael Neff, and Jonas Beskow. 2017. Mimebot—investigating the expressibility of non-verbal communication across agent embodiments. *ACM Transactions on Applied Perception (TAP)* 14, 4 (2017), 1–13.

[2] Roberto Bresin, Anna de Witt, Stefano Papetti, Marco Civolani, and Federico Fontana. 2010. Expressive sonification of footstep sounds. *Proceedings of ISon* 2010 (2010), 51–54.

[3] Roberto Bresin, Thomas Hermann, and Andy Hunt. 2012. Interactive sonification.

[4] Emma Frid, Roberto Bresin, and Simon Alexanderson. 2018. Perception of Mechanical Sounds Inherent to Expressive Gestures of a NAO Robot-Implications for Movement Sonification of Humanoids. *Sound and Music Computing* (2018).

[5] B.L. Giordano, H. Egermann, and Roberto Bresin. 2014. The production and perception of emotionally expressive walking sounds : Similarities between musical performance and everyday motor activity. *PLoS ONE* 9, 12 (2014), e115587–. https://doi.org/10.1371/journal.pone.0115587 QC 20150126.

[6] Andy Hunt and Thomas Hermann. 2011. Interactive sonification. In *The sonification handbook*.

[7] Eun-Sook Jee, Yong-Jeon Jeong, Chong Hui Kim, and Hisato Kobayashi. 2010. Sound design for emotion and intention expression of socially interactive robots. *Intelligent Service Robotics* 3, 3 (2010), 199–206.

[8] Gregory Kramer, BN Walker, Terri Bonebright, Perry Cook, J Flowers, Nadine Miner, John Neuhoff, R Bargar, S Barrass, J Berger, et al. 1999. The sonification report: Status of the field and research agenda. report prepared for the national science foundation by members of the international community for auditory display. *International Community for Auditory Display (ICAD), Santa Fe, NM* (1999).

[9] Adrian Benigno Latupeirissa, Emma Frid, and Roberto Bresin. 2019. Sonic characteristics of robots in films. In *Sound and Music Computing Conference*. 1–6.

[10] Adrian Benigno Latupeirissa, Claudio Panariello, and Roberto Bresin. 2020. Exploring emotion perception in sonic HRI. In *Sound and Music Computing Conference, Torino, 24-26 June 2020*. Zenodo, 434–441.

[11] Jonas Löwgren. 2007. Fluency as an experiential quality in augmented spaces. *International Journal of Design* 1, 3 (2007), 1–10.

[12] Claudio Panariello, Mattias Sköld, Emma Frid, and Roberto Bresin. 2019. From vocal sketching to sound models by means of a sound-based musical transcription system. In *Sound and Music Computing Conference, Universidad de Malaga Malaga, Spain, May 28-31, 2019*. 1–7.

[13] Davide Rocchesso, Guillaume Lemaitre, Patrick Susini, Sten Ternström, and Patrick Boussard. 2015. Sketching Sound with Voice and Gesture. *interactions* 22, 1 (2015), 38–41.

[14] Davide Rocchesso, Stefania Serafin, and Michal Rinott. 2013. Pedagogical Approaches and Methods. *Sonic Interaction Design* (2013), 125–150.

[15] Nathan Shedroff and Chris Noessel. 2012. Make it so: Learning from sci-fi interfaces. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 7–8.

[16] Sköld, M. 2018. Combining Sound- and Pitch-Based Notation for Teaching and Composition. In *TENOR'18 – Fourth International Conference on Technologies for Music Notation and Representation*. 1–6.

[17] René Tünnermann, Jan Hammerschmidt, and Thomas Hermann. 2013. Blended Sonification: Sonification for Casual Interaction. In *ICAD 2013-Proceedings of the International Conference on Auditory Display*.

[18] Aravind Elanjimattathil Vijayan, Simon Alexanderson, Jonas Beskow, and Iolanda Leite. 2018. Using Constrained Optimization for Real-Time Synchronization of Verbal and Nonverbal Robot Behavior. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1955–1961.

[2] https://consequentialrobotics.com/miro-beta