2020 29th IEEE International Conference on
Robot and Human Interactive Communication (RO-MAN)
August 31 - September 4, 2020 - Virtual Conference

Robot Gesture Sonification to Enhance Awareness of Robot Status and Enjoyment of Interaction

Lisa Zahray[1], Richard Savery[1], Liana Syrkett[1], and Gil Weinberg[1]

Abstract— **We present a divergent approach to robotic sonification with the goal of improving the quality and safety of human-robot interactions. Sonification (turning data into sound) has been underutilized in robotics, and has broad potential to convey robotic movement and intentions to users without requiring visual engagement. We design and evaluate six different sonifications of movements for a robot with four degrees of freedom. Our sonification techniques include a direct mapping from each degree of freedom to pitch and timbre changes, emotion-based sound mappings, and velocity-based mappings using different types of sounds such as motors and music. We evaluate these sonifications using metrics for ease of use, enjoyment/appeal, and conveyance of movement information. Based on our results, we make recommendations to inform decisions for future robot sonification design. We suggest that when using sonification to improve safety of human-robot collaboration, it is necessary not only to convey sufficient information about movements, but also to convey that information in a pleasing and even social way to to enhance the human-robot relationship.**

I. INTRODUCTION

With the increasing role of robots in the workplace, homes, and society as a whole, safety when interacting with robots is of great importance [1]. Safety is important for working with industrial robots where there is potential for serious injury, but it is also relevant for service robots in the home [2]. Robotic safety research often focuses on physical aspects of the robot, such as improving its reliability and sensing capabilities. However, a human's awareness of a robot's actions is also important, as perceived safety when working with a robot has been shown to improve willingness to interact with the robot in the future [3].

Sonification is the study of using "nonspeech audio to convey information... for purposes of facilitating communication or interpretation" [4]. In this paper, we explore sonifying a robot's actions for the purpose of improving critical HRI metrics [5] for effective collaboration and utilization, focusing primarily on perceived safety through movement information conveyance, and likeability through enjoyment/appeal ratings. While human gesture sonification has been widely studied, there is minimal existing research on robot sonification. In this paper, we expand on that research by evaluating several robot sonification techniques on criteria related to enjoyment, ease of use, and conveyance of movement information. We then propose approaches for how our results can be used to improve the quality and safety of human-robot interactions.

*This work was not supported by any organization

[1] The authors are with Georgia Tech Center for Music Technology, Atlanta, GA 30332, USA
lzahray3@gatech.edu, rsavery3@gatech.edu, liana.syrkett@gatech.edu, gilw@gatech.edu

II. RELATED WORK

A. Safety in Human Robot Interaction

Due to the increasing presence of robots in work and social situations, significant research is being conducted for robotic safety [6]. This research has largely focused on robotic performance, such as collision detection and reliability [7]. A supporting approach is giving the human feedback from the robot, to allow them to anticipate and follow the robot's path. One example is multi-modal research where human operators are given audio and haptic feedback [8]. For social robotics, carefully constructed multi-modal feedback or audio based feedback can increase common HRI metrics, such as trust [9]. Additionally the perception of clear feedback and the perception of safety affects a person when interacting with a robot [10], extending from physical safety to broader privacy and psychological concerns [11]. While different forms of feedback allow for better safety through human-awareness and perception, there are also many possibilities associated with social robotics for entertainment and increased collaboration opportunities [12]. Aside from natural language and speech, audio has received limited attention, with most focus instead on gesture and physical features [13].

B. Gesture and Movement Sonification

Gesture sonification has been used for a variety of purposes. [14] investigated head gesture sonification to support social interaction for visually impaired persons. In this case, sonification was used to convey information for improving human-human interactions. [15] sonified 2-d gestures in order to teach visually impaired persons to perform the gesture, finding that pitch best mapped to vertical movements while stereo panning best mapped to horizontal movements. In this paper, we further investigate those mappings with our pitch and timbre sonification. Many dance sonification projects have also been conducted, such as real-time audio based on dancers' movements [16]. This is an example of gesture sonification used primarily for entertainment. Sonification of physical gestures has additionally been used widely in sports such as optimizing performance in rowing [17].

C. Robotic Sonification

There has been minimal sonification research in relation to robotics. Moroni and Manzolli used evolutionary algorithms to generate musical compositions that then control robots [18], although this is only loosely connected to the process of sonification. Zhang et al. [19] studied robotic sonification in relation to emotions for children with Autism Spectrum Disorder (ASD). In their work, they based their emotional states on the research of [20]. In our proposed system, we also use the emotional descriptor mappings that [20] has created as the basis of our emotion-based sonification. There have also been studies on the mechanical sound of robots, or the sounds that robots inherently make when moving through processes [21]. We explore this idea in one of our sonification techniques, which uses motor sounds.

III. Motivation

Sonification of gestures has been shown to be effective for improving social interactions, entertainment, and conveying movement information. This suggests that sonification could also be useful for HRI, both for improving the social aspect of interactions, as well as conveying movement information to improve perceived safety. In this paper, we take ideas from previous work in gesture and robot sonification to create several robot sonification types, ranging from information-conveying mappings to mappings involving entertainment and emotion. We then compare and evaluate these sonifications. Our research questions can be summarized as:

1) What robot sonification techniques are the most effective for conveying movement information?
2) What robot sonification techniques do users find most engaging and enjoyable?

IV. Sonification Techniques

We categorized our sonification techniques within Walker and Nees' theory of sonification [22]. They list four functions of sonification: alarm, status, art and entertainment, and data exploration. This work primarily focuses on status showing and entertainment possibilities, aiming to cross safety applications with entertainment approaches that could be used in social robotics.

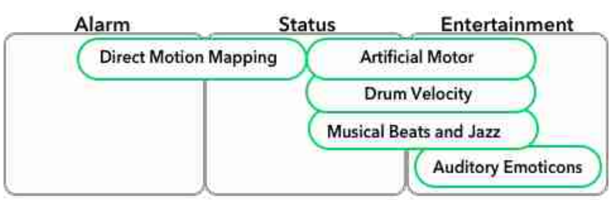

Fig. 1. Implemented Robotic Categorization

We use the robot Shimon for our sonifications. His four degrees of freedom are shown in Figure 2. Videos of the sonifications are available here[1].

A. Direct Pitch and Timbre

Our first technique directly maps sounds to the movements of each of Shimon's degrees of freedom (DOF). We use one high-pitched sawtooth wave for Shimon's head DOF's, and a low-pitched sawtooth wave for Shimon's body DOF's. For each waveform, pitch is mapped to the vertical DOF, where a higher position maps to a higher pitch. A low-pass filter is mapped to the horizontal DOF, where the cutoff frequency is higher the further the head moves to the right. This creates a more muted sound on the left, and a more buzzy sound on the right. During movements, we linearly interpolate the pitch and cutoff frequency to smoothly approach their target value, reaching it at the same time the corresponding DOF reaches its target position. We implemented these mappings using the Max/MSP programming language[2].

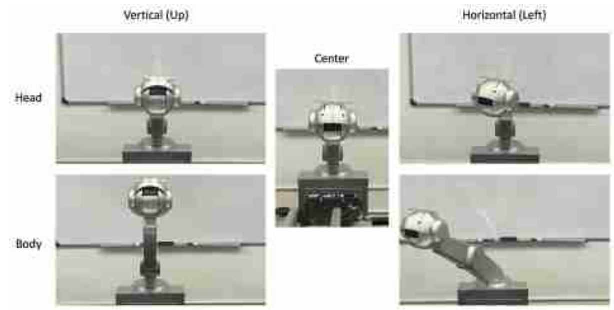

Fig. 2. Shimon's 4 DOFs

B. Auditory Emoticons

Our next technique maps emotional descriptor-based sounds to Shimon's movements. Basing our mapping on the work of [20], we chose seven auditory emoticons that have already shown to correspond to people's emotional states. We chose auditory icons over earcons because we wanted to create a juxtaposition of naturally occurring sounds along with the artificial sounds we created. We then mapped these emotional descriptors to movements in an effort to enhance participants' perceived emotions. For example, we mapped the angry sounds to sharp, quick movements.

C. Artificial Motor Sounds

Shimon's movements and motor sounds are hardly audible in day to day use, with head movements close to silent. We developed an approach that uses real sampled motor sounds from a sampled electric motor and mapped these to Shimon's movements. For this sonification, we kept the mapping as simple as possible, using only the speed of Shimon with no positional information. First, we created a arbitrary normalized speed between 0.0 and 1.0 for each of Shimon's four degrees of freedom, leading to a total speed between 0.0 and 4.0. This is then mapped to a time-stretching algorithm for the motor sounds, with pitch coupled, meaning pitch increases as the speed increases. The speed range is linearly mapped to the audio speed with 0.0 equalling 40% playback speed, and 4.0 equaling 300% playback speed. This mapping was subjectively chosen after internal testing of different speed mappings.

[1] http://www.richardsavery.com/robotsonification

[2] https://cycling74.com/

D. Percussion Sounds

The next sonification builds on the mappings used by artificial motor sounds, using similar mapping for Shimon's speed. In this case, 0.0 to 4.0 is mapped to 40% and 1,000% speed respectively, with pitch decoupled. The samples are 1,000 ms long clips of a snare drum, cymbal and floor tom. In addition to the speed of the playback, the percussive sonification includes the horizontal body movements, which is Shimon's largest range of movement. Movements to the left increase the volume of the cymbal, while movements to right increase the volume of the floor tom, with both direction reducing the volume of the snare. When Shimon is at the extreme of any position, only a single sound is heard.

E. Musical Loops Based Sonifications

This sonification maps rhythmic musical loops to Shimon's movement. Similarly to Drums and Motor sounds, this sonification also uses Shimon's movement speed, mapped to the tempo; 0.0 to 4.0 is mapped to 80% to 200% respectively, or in this case, 80 beats per minute to 200 beats per minute. A lower range of speed variations was chosen, as subtle variations in musical tempo were much easier to detect than for the motor or drum speed. In addition to speed, a musical instrument is mapped to each of Shimon's moving components. A drumkit playing a groove is mapped to horizontal body movements, a bass to horizontal head and vertical body, and guitar/chords to vertical head. When Shimon moves, the corresponding instrument sound plays, and when there is no movement each instrument remains silent. The loops are synchronized and faded in and out, maintaining the structure of the loop independently of the movement. Two separate genres were used for samples, one rock-based (called Beats) and the other Jazz.

V. Evaluation Procedure

Gesture Name	Description	Duration (s)
Worm	Head & body move up and down while head & body move right and then left	8.2
Down Across	Head & body move down & left, then head & body move right	8.3
Side Speedup	Body moves slowly left, then head & body move quickly left	3.8
Nodding	Head nods up & down and looks right to left, while body moves slowly up then down	7.5
Rise Right	Head & body move slowly up and right	4.9
Sharp Moves	Series of fast, isolated movements	4.2
Long	Longer series of various medium-speed movements	16.7

TABLE I

Gesture Descriptions and Details

We evaluated all sonification techniques in a listening test using a total of 30 participants divided among 10 groups of 2 to 5 participants each. Each group watched all gestures performed live on the robot, operated by one researcher who was present in the room. Sounds were produced by a speaker placed on the floor underneath the robot. The participants first evaluated the Direct Pitch and Timbre sonification, followed by Auditory Emoticons, since these sonification methods used unique surveys and procedures as described in the following subsections V-A and V-B. The remaining four sonification technique evaluations were randomly ordered.

All sonification techniques utilized the same set of gestures in their evaluation processes. These are listed in Table I, along with the corresponding emotional descriptor mappings that were used for the Auditory Emoticons sonification technique. The list includes three short gestures (less than 5 seconds), three medium-length gestures (between 7 and 9 seconds), and one long gesture (16.7 seconds).

A. Direct Pitch and Timbre: Movement Identification

This experiment evaluates participants' ability to identify details of the robot's motion from the Pitch and Timbre sonification without looking at the robot. We explore the following three sub-research questions of Research Question 1, evaluating conveyance of movement information:

1.A) Is the sonification intuitive: Will the participants score better than random chance before seeing how the sounds are mapped to the robot movements?
1.B) Is the sonification learnable: Will the participants score better after watching the robot move with the corresponding sounds?
1.C) Is there any difference in performance between horizontal movements (mapped to timbre) and vertical movements (mapped to pitch)?

To answer these research questions, we evaluated three different categories of motion: sequential, simultaneous, and gestures. Participants were first shown the individual movements that the robot could perform (consisting of all body part, direction, and speed combinations) without sound. The robot was then turned off, and the participants were asked to identify movement details for each motion category based solely on listening to the sounds. First, participants listened to 2 sequential movements and were asked to identify the body part (head or body), the direction (up, down, left, or right), and the speed (fast or slow) of each movement. This process was repeated for 3 sets of sequential movements. Participants next listened to 2 simultaneous movements and were asked to identify the same movement properties. They were told they list the two simultaneous movements in any order. This process was repeated for 3 sets of simultaneous movements. Finally, participants listened to 3 separate gestures and were asked to provide a description of what they believed the robot was doing during each gesture.

Following this process, the participants observed the robot moving while listening to the sonification for 3 minutes. We refer to this as "training" participants on the sonification. Participants were then asked to repeat the same movement identification test. The movements used before and after this training process were all randomized for each group. The 6 gestures used were all gestures in Table I, excluding "Long".

The gestures were randomly assigned to the before or after test for each group.

B. Emotional Descriptor Identification

For the Auditory Emoticons sonification evaluation, we evaluated if the participants were able to determine which emotional descriptor corresponded to the sonification. For each sonification, participants had the option of choosing from the emotional descriptors presented in Table I. Furthermore, we asked the participants if the sonification activated their other senses and / or their emotions. These questions were formed using a 7 point Likert scale [23] where 1 represented "Not at all/to a low degree" and 7 represented "Very often/to a high degree."

C. BUZZ Scale

For each sonification technique, the participants also completed the Buzz Scale [24]. The Buzz Scale consists of 11 questions evaluating the audio user experience. The questions were developed from other usability metrics and combine to give an ease of use rating and an enjoyment/appeal rating. For sonification practice, ease of use refers to the ability for users to understand the meaning of the sounds and their relation to the data.

VI. Results

A. Direct Pitch and Timbre

This section presents our evaluation of the "before" and "after" training movement identification surveys. Data from one 2-person group survey was excluded due to a bug that resulted in unusable results. Therefore, the overall data used for this analysis included 28 participants across 9 groups.

1) Sequential Movements: We calculated the percentage of correct answers each participant gave for each of the following movement categories, both before and after training: body part (head vs. body), main direction (vertical vs. horizontal), specific direction (right, left, up, or down), and speed (fast or slow). The expected value for each movement category with random guessing is 0.5, except for the specific direction category which is 0.25. Figure 3 shows these percentages compared against the expected values, as well as the overall percentage of correct answers across all movement categories (expected value with random guessing 0.4375).

To test Research Question 1.A (intuitiveness), we performed a 1-sample Hotelling's T-squared test on participants' "before" percentages, comparing against their expected values with random guessing. We chose this test because our data is multivariate with four separate movement categories. The resulting p-value is 0.0205 which is less than the alpha of 0.05, supporting that at least one "before" movement category has a different accuracy than expected with random guessing. To find which movement categories were different, we next performed a 1-sample, 2-tailed t-test on each category. We found one significant p-value for the main direction category; the p value was 0.00108, which is less than the alpha of 0.05. This supports that participants were able to identify whether the movement was vertical or horizontal with better accuracy than random chance before being trained on the sonification.

To test Research Question 1.B (learnability), we first performed a paired Hotellings t-squared test between participants' "before" and "after" percentages on the four movement categories. The p value is 3.94e-04, which is less than the alpha of 0.05, supporting that there is at least one difference between the "before" and "after" percentages. We next performed individual 2-tailed, paired t-tests on participants' "before" and "after" percentages for each of the four movement categories. Three p-values were less than the alpha of 0.05 and therefore significant: main direction ($p = 0.0379$), specific direction ($p = 7.04e-06$), and speed ($p = 0.0212$). The data support that after training, participants performed better than before training on these three movement categories, but did not perform better on body part identification.

To test Research Question 1.C (horizontal vs. vertical), we performed four Chi-Square tests (one for each movement category) to determine whether the frequencies of correct vs. incorrect answers for vertical movements differed from those of horizontal movements. The Chi-Square Test was performed on the "after" data and produced the following p-values: body part 0.0118, main direction 0.387, specific direction 0.132, and speed 0.963. Figure 4 shows a comparison of overall correct answer percentages for vertical vs. horizontal movements. Body part is the only significant movement type with a p-value less than the alpha of 0.05. The results support that participants performed better at identifying the body part for vertical movements than for horizontal movements. Interestingly, this was the only movement category that participants did not improve on after training. This makes sense, as participants' difficulty in identifying the body part for horizontal movements was detrimental to their overall accuracy in that category.

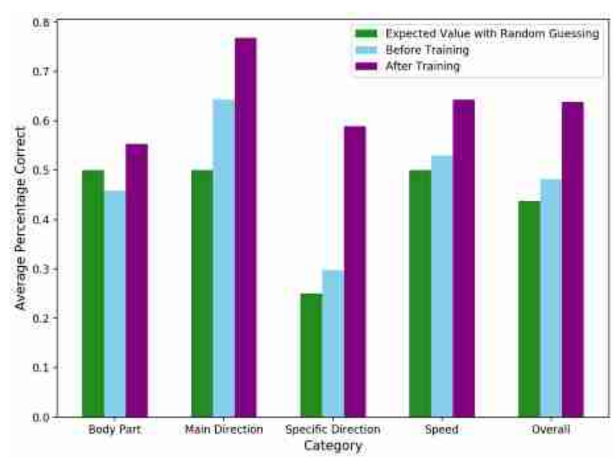

Fig. 3. Sequential Movement Accuracy

2) Simultaneous Movements: We evaluated participants' accuracy in identifying simultaneous movements to further test Direct Mapping Research Questions 1.A and 1.B. Because each pair of movements happens simultaneously, we

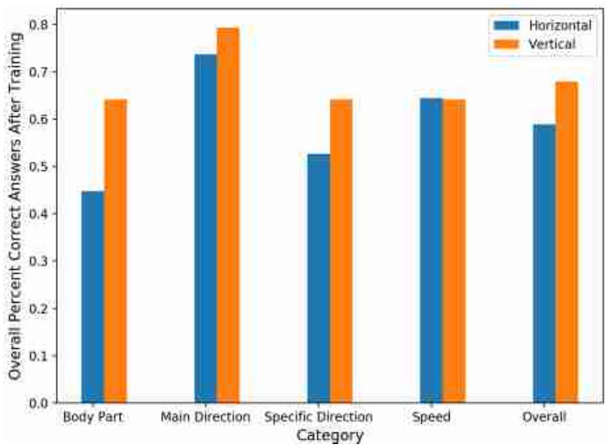

Fig. 4. Horizontal vs. vertical accuracies after training

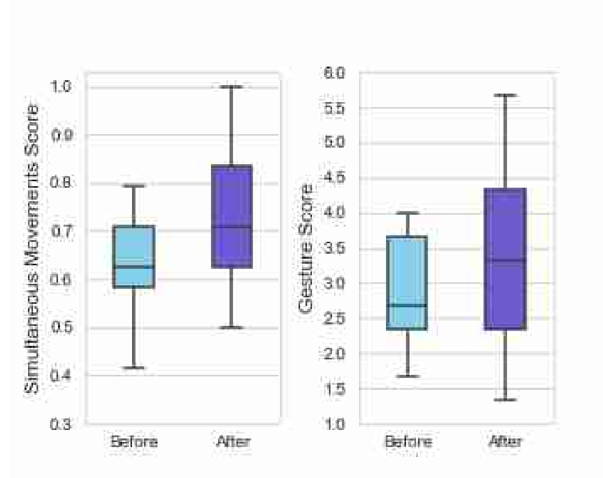

Fig. 5. Simultaneous Movement (left) and Gesture (right) Scores

must define a way to evaluate participants' accuracy without a known mapping between their two reported movements and the two actual movements. To address this, we define a participants' simultaneous movement score as the maximum possible percentage of correct movement answers by trying both possible assignments of the two movements. The expected value of this score for random (but valid) guessing is 0.611, calculated by simulating all possible valid guess/actual movement pairs and taking the average score. By "valid", we mean that the same body part cannot simultaneously move in the same main direction (i.e., the head cannot simultaneously go right and right or right and left).

The results for the simultaneous movement scores before and after training are shown in Figure 5. We first performed a 1-sample, 2-tailed t-test on the "before" scores to test whether they were different than random chance. The p-value was 0.276 which is not significant. We then performed a paired 2-tailed t-test between participants' "before" and "after" scores. The p-value was significant with a value of 2.43e-04. The data support that participants did not perform better than random chance before training, but improved after training.

3) Gestures: Gestures were the final movement type we used to evaluate Research Questions 1.A and 1.B. All gesture descriptions were matched with their correct gesture, and the order was randomized so that the researcher scoring the answers did not know which descriptions came from the "before" or "after" survey. Descriptions were rated from 1 to 7, with 7 corresponding to perfect identification of all movements in the gesture, and 1 corresponding to completely incorrect movement identification. The results are shown in Figure 5. We assigned each participant a "before" and "after" gesture score by taking the average of their corresponding description ratings. We performed a paired 2-tailed t-test on the "before" and "after" gesture scores for each participant. The p-value is 0.0294 which is less than the alpha of 0.05, supporting that participants were able to provide better gesture movement descriptions after training. However, the average gesture score after training is still fairly low, indicating that many participants were still missing or incorrectly identifying key movements in the gestures.

We also analyzed the descriptions qualitatively. Participants could often tell how much overall movement was occurring both before and after training, but struggled to correctly identify the particular directions, especially for simultaneous movements. Some horizontal movements were occasionally missed entirely or mistaken for vertical movements. Vertical movements were usually correctly identified, even in many "before" descriptions. Likely due to the nature of the previous survey questions, participants usually phrased their responses in terms of specific body parts and directions. However, several comments before training described gestures in a more general way, such as "The robot was scanning around the place" and "Robot waking up, turning on his head and body." These types of responses tended to score low due to lack of specificity, but were in fact fairly accurate portrayals of the intention behind the gestures' designs.

B. Auditory Emoticons

The results of the Auditory Emoticons sonification evaluation are shown in Table II. This table shows the percentage of participants that chose the goal emotional descriptor for the corresponding sonification, as well as the most common emotional descriptor chosen for the sonification. Our results show that the most commonly chosen emotional descriptor almost always matched the goal emotional descriptor for each gesture. The only gesture where this was not the case was the rise right gesture. The most commonly chosen emotional descriptor was modern, but the actual emotional descriptor being demonstrated was dreamy. For all of the gestures, over 30% of the participants were able to accurately match the emotional descriptor to the gesture. For the two Likert scale questions shown in Figure 6, we found that many participants answered in the neutral 4 area. However, more participants that leaned toward the higher values of the scale than the lower values.

Gesture Name	Sound	Goal Descriptor	Goal Descriptor Percentage	Most Commonly Chosen
Worm	Breeze	Calm	62.5%	Calm
Down Across	Sigh	Boring	50%	Boring
Side Speedup	Clock	Simple	34.38%	Simple
Nodding	Cheering	Lively	90.63%	Lively
Rise Right	Pulsing	Dreamy	31.25%	Modern (with 37.5%)
Sharp Moves	Honking	Angry	59.38%	Angry
Long	Typing	Modern	31.25%	Modern

TABLE II
RESULTS OF EMOTIONAL DESCRIPTOR TESTING

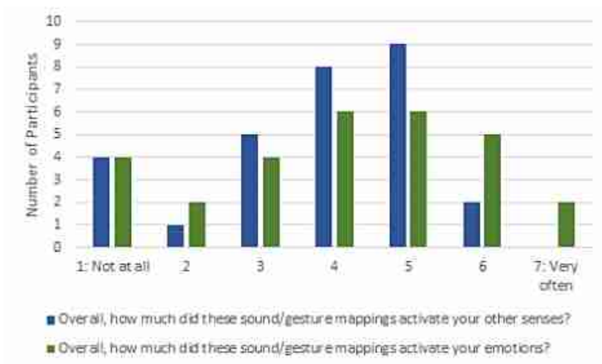

Fig. 6. Responses to Auditory Emoticon Sonification

C. BUZZ Scale Scores

After collecting the Buzz scores, we first inverted the negative results. For each participant, we then combined 6 of the scales to create an ease of use score, 5 to create an enjoyment/appeal result and all 11 for a composite result. Figure 7 shows the results from the combinations. We then conducted a repeated measure ANOVA for ease of use and enjoyment/appeal. Figure 8 and Figure 9 show the results.

D. Feedback from Comments

1) Direct Pitch and Timbre: Many respondents mentioned how the sounds were unpleasant, specifically the high pitch sound. One participant mentioned that the sounds were distracting to the task because they were thinking about the harshness of the sounds instead of focusing on where the robot was moving. Other participants mentioned how the up/down mappings were easier to determine whereas the left/right ones were more difficult. Another participant mentioned that after watching Shimon move, they were able to distinguish the left/right pair, but without seeing the robot move, the task became difficult again. While most participants mentioned that higher pitch mapped to upward movement made sense, a different participant mentioned that the high pitch could be mapped with downward movement and still have the same effect.

2) Auditory Emoticons: In general, the participants enjoyed the sounds. Many participants mentioned that the sounds were generally easy to map to the emotional descriptor, with the exception of the clock. One participant mentioned how being familiar with the sounds made it easier to create these emotional descriptor mappings. However, there was general consensus that the mappings of the sounds to Shimon's gestures were arbitrary and difficult to correlate.

3) Motor: Many of the participants mentioned that it was easy to tell Shimon's speed based on the speed of the motor sound. However, they also mentioned that the other movements besides the tempo-based ones were difficult to distinguish. Participants also mentioned that the sounds themselves did not have as distinguishable differences between the ones corresponding to the head and body. Many of the participants felt confused, and one participant mentioned that Shimon looked confused by the sounds as well.

4) Drums: A few participants mentioned that the sounds were not pleasant to listen to, with one describing them as tense and abrupt, and another participant mentioning that the sounds made them "almost angry." In general, the participants had a difficult time creating the associations between the gestures and the sounds. One participant mentioned the sounds not being different enough to make them distinguishable for the movements. A few participants did mention that they were able to map the gestures, but for the most part, the participants felt that they were random.

5) Music Based: Beats: There was a general consensus that having multiple actions happening at once made it difficult for participants to decipher which gesture corresponded to the sound. Some participants said they were able to correctly identify some gestures, but not all of them. One participant said the bass sound in particular was difficult to map. Another participant mentioned that after a few more listens, they would have been able to distinguish where Shimon would be moving with their eyes closed. Most participants also indicated that in comparison to the other sonification categories, this one was easier to identify the connection between the sound and the movements. While a few participants did not feel that the movements matched with the music, the large majority of participants mentioned how they enjoyed the sounds and that the movements matched them well. One participant indicated that they could see Shimon performing these gestures on stage.

6) Music Based: Jazz: In general, the participants enjoyed the sounds, describing them as entertaining. One participant mentioned that Shimon seemed to enjoy the music too. However, overall participants found that there was not a strong correlation between the sound itself and the gesture. A few participants felt that they could not distinguish what was occurring until Shimon's head and body moved independently. They also mentioned that there were many things to keep track of, making the task difficult to accomplish.

VII. DISCUSSION AND FUTURE WORK

In this section, we discuss how our results can help inform design decisions for creating robot sonifications that improve the quality of human-robot interactions.

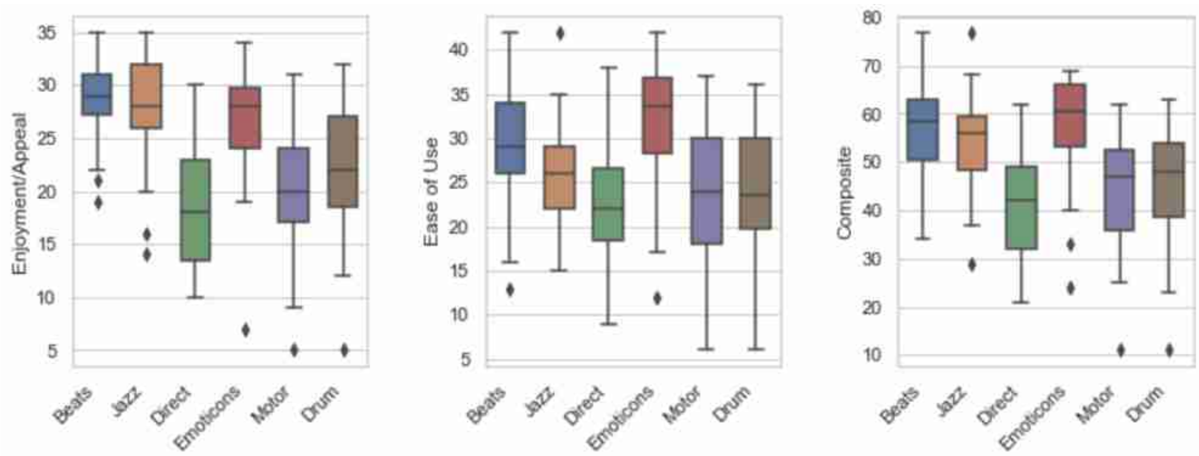

Fig. 7. Box Plots: Ease of Use Results (out of 42), Enjoyment (out of 35), Composite (out of 77)

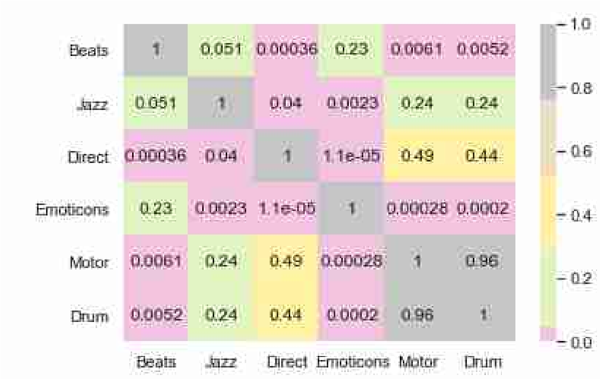

Fig. 8. Ease of Use Heat Map of P-values (Pink is significant)

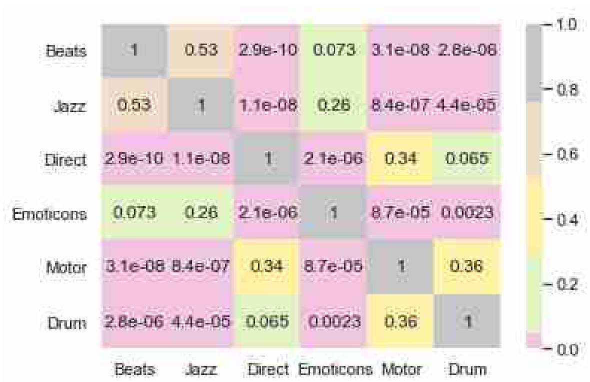

Fig. 9. Enjoyment/Appeal Heat Map of P-values (Pink is significant)

We first discuss successful and unsuccessful features of the Direct Pitch and Timbre sonification for movement identification. Participants were able to improve on simultaneous movements, indicating that different sound changes simultaneously produced by our mappings were clear enough to be identifiable. However, participants' inability to improve on body part identification indicates that our approach of assigning two differently-pitched waveforms to each body part did not create a clear enough distinction in all cases. This was especially true for body part identification during timbre changes (horizontal movements). Participants naturally associated vertical sounds with pitch changes and horizontal sounds with timbre changes better than random chance, indicating an intuitiveness of these mappings. We find it surprising that participants did not naturally associate upwards and downwards motion with upwards and downwards changes in pitch respectively, with some participants guessing the reverse. Overall, vertical robot movements mapped to pitch was intuitive and identifiable, while timbre changes may be too subtle for the purpose of robot movement identification. Finally, it may be possible to isolate changes to two distinctly-pitched waveforms if only modifying pitch, but not when using timbre changes.

The Buzz score results indicate that participants preferred the Beats, Jazz, and Emoticons sonifications over Direct, Motor, and Drum. These three preferred sonifications used music and emotional descriptor-based sound recordings, which were reported as being more pleasant and familiar than filtered sine waves, motor sounds, and abrupt percussive hits. It is interesting that these three sonifications had higher ratings not just for enjoyment/appeal, but also for ease of use. Two examples of criteria used for this metric are "it was easy to match these sounds to their meanings" and "the sounds are relatable to their ideas". For emoticons, participants likely interpreted these criteria as relating the sounds to their emotions, rather than the robot movements. Beats and Jazz scored higher than Direct for ease of use, even though they do not provide enough information to fully specify the robot's location. This indicates that velocity-based mappings and/or selective fading of audio tracks (used by Beats and Jazz) may be more effective for indicating a robot's status than position-based mappings and pitch/timbre changes (used by Direct). It would be worth exploring different mapping types with movement identification tests for future work.

We believe that preference for the music and emotional descriptor-based sonifications shows that to help a person feel as if they understand a robot sonification, the sounds should be enjoyable and allow the person to relate to the robot. These qualities may also better incentivize someone to spend more time with the robot to learn the sonification mappings. When using sounds to improve safety in HRI, it is not only important to have a clear mapping, but also for that mapping to be pleasant, inviting, and even social; perceived robotic safety is not just about conveying the robot's exact position, but also about improving the relationship between humans and robots so people feel more comfortable.

Future work could include the design of new robot sonifications informed by these results, and evaluate them for movement identification ability as well as HRI metrics. This could include further investigation into what aspects of sound people naturally associate with different movements. Another option for future work is evaluating what sonification types work best for conveying a robot's intention, and whether different sonifications can convey different interpretations of the same gesture. In our Emoticons sonification, we mapped sounds to gestures subjectively. Future work could also include testing whether mapping sounds to different gestures could change participants' choice of the emotional descriptor. We would also like to explore different approaches to an emotional descriptor-based mapping that are more dependent on the robot's movements.

VIII. Conclusion

This paper presented the design and evaluation of several different sonification techniques for robot movements for the purposes of improving perceived safety and likeability in HRI. For our direct mapping, before seeing how the sonification was mapped, participants did better than random chance on identifying vertical vs. horizontal movement, but no other specific movement aspects. Participants were able to improve at identifying sequential, simultaneous, and gestural movements after observing the sonification with the robot; however, they performed worse for horizontal movements (timbre) than vertical (pitch). For our emotional descriptor-based sonification, participants were generally able to correctly match the intended emotional descriptor with its sonification. However, this sonification provides little information about the robot's movements. We evaluated all sonifications using the Buzz Scale, and found that participants favored the emotional descriptor-based sonification and the music-based sonifications over the other techniques. Based on these results, we propose that sonification for HRI should not only convey movement information, but also be enjoyable and familiar in order to enhance the human-robot relationship.

References

[1] S. Robla-Gómez, V. M. Becerra, J. R. Llata, E. Gonzalez-Sarabia, C. Torre-Ferrero, and J. Perez-Oria, "Working together: A review on safe human-robot collaboration in industrial environments," *IEEE Access*, vol. 5, pp. 26 754–26 773, 2017.

[2] T. S. Tadele, T. de Vries, and S. Stramigioli, "The safety of domestic robotics: A survey of various safety-related publications," *IEEE Robotics Automation Magazine*, vol. 21, no. 3, pp. 134–142, 2014.

[3] S. You, J.-H. Kim, S. Lee, V. Kamat, and L. P. Robert Jr, "Enhancing perceived safety in human–robot collaborative construction using immersive virtual environments," *Automation in Construction*, vol. 96, pp. 161–170, 2018.

[4] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. Flowers, N. Miner, J. Neuhoff, R. Bargar, S. Barrass, J. Berger, *et al.*, "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.

[5] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[6] G. Herrmann and C. Melhuish, "Towards safety in human robot interaction," *International Journal of Social Robotics*, vol. 2, no. 3, pp. 217–219, 2010.

[7] B. S. Dhillon and A. Fashandi, "Safety and reliability assessment techniques in robotics," *Robotica*, vol. 15, no. 6, pp. 701–708, 1997.

[8] S. Papanastasiou, N. Kousi, P. Karagiannis, C. Gkournelos, A. Papavasileiou, K. Dimoulas, K. Baris, S. Koukas, G. Michalos, and S. Makris, "Towards seamless human robot collaboration: integrating multimodal interaction," *The International Journal of Advanced Manufacturing Technology*, vol. 105, no. 9, pp. 3881–3897, 2019.

[9] R. Savery, R. Rose, and G. Weinberg, "Establishing human-robot trust through music-driven robotic emotion prosody and gesture," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2019, pp. 1–7.

[10] J. Young, R. Hawkins, E. Sharlin, and T. Igarashi, "Toward acceptable domestic robots: Applying insights from social psychology," *International Journal of Social Robotics*, vol. 1, no. 95, 2009.

[11] T. Denning, C. Matuszek, K. Koscher, J. R. Smith, and T. Kohno, "A spotlight on security and privacy risks with future household robots: Attacks and lessons," in *Proceedings of the 11th International Conference on Ubiquitous Computing*, ser. UbiComp '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 105–114.

[12] C. Breazeal, K. Dautenhahn, and T. Kanda, "Social robotics," in *Springer handbook of robotics*. Springer, 2016, pp. 1935–1972.

[13] S. Saunderson and G. Nejat, "How robots influence humans: A survey of nonverbal communication in social human–robot interaction," *International Journal of Social Robotics*, vol. 11, no. 4, pp. 575–608, 2019.

[14] T. Hermann, A. Neumann, and S. Zehe, "Head gesture sonification for supporting social interaction," in *Proceedings of the 7th Audio Mostly Conference: A Conference on Interaction with Sound*, 2012, pp. 82–89.

[15] U. Oh, S. K. Kane, and L. Findlater, "Follow that sound: Using sonification and corrective verbal feedback to teach touchscreen gestures," in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ser. ASSETS '13. New York, NY, USA: Association for Computing Machinery, 2013.

[16] S. Landry and M. Jeon, "Participatory design research methodologies: A case study in dancer sonification," in *The 23rd International Conference on Auditory Display (ICAD 2017)*, State College, PA, USA, 2017.

[17] N. Schaffert, K. Mattes, and A. O. Effenberg, "A sound design for the purposes of movement optimisation in elite sport (using the example of rowing)." Georgia Institute of Technology, 2009.

[18] A. Moroni and J. Manzolli, "From evolutionary composition to robotic sonification," in *European Conference on the Applications of Evolutionary Computation*. Springer, 2010, pp. 401–410.

[19] R. Zhang, J. Barnes, J. Ryan, M. Jeon, C. H. Park, and A. Howard, "Musical robots for children with asd using a client-server architecture," in *International Conference on Auditory Display*, 2016.

[20] J. Sterkenburg, M. Jeon, and C. Plummer, "Auditory emoticons: Iterative design and acoustic characteristics of emotional auditory icons and earcons," in *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*, M. Kurosu, Ed. Cham: Springer International Publishing, 2014, pp. 633–640.

[21] E. Frid, R. Bresin, and S. Alexanderson, "Perception of mechanical sounds inherent to expressive gestures of a nao robot - implications for movement sonification of humanoids," 07 2018.

[22] B. N. Walker and M. A. Nees, "Theory of sonification," *The sonification handbook*, pp. 9–39, 2011.

[23] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 1–55, 1932.

[24] B. J. Tomlinson, B. E. Noah, and B. N. Walker, "Buzz: An auditory interface user experience scale," in *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, pp. 1–6.