Research Article

Richard Savery*, Lisa Zahray, and Gil Weinberg

Emotional musical prosody for the enhancement of trust: Audio design for robotic arm communication

https://doi.org/10.1515/pjbr-2021-0033
received March 1, 2021; accepted November 19, 2021

Abstract: As robotic arms become prevalent in industry, it is crucial to improve levels of trust from human collaborators. Low levels of trust in human–robot interaction can reduce overall performance and prevent full robot utilization. We investigated the potential benefits of using emotional musical prosody (EMP) to allow the robot to respond emotionally to the user's actions. We define EMP as musical phrases inspired by speech-based prosody used to display emotion. We tested participants' responses to interacting with a virtual robot arm and a virtual humanoid that acted as a decision agent, helping participants select the next number in a sequence. We compared results from three versions of the application in a between-group experiment, where the robot presented different emotional reactions to the user's input depending on whether the user agreed with the robot and whether the user's choice was correct. One version used EMP audio phrases selected from our dataset of singer improvisations, the second version used audio consisting of a single pitch randomly assigned to each emotion, and the final version used no audio, only gestures. In each version, the robot reacted with emotional gestures. Participants completed a trust survey following the interaction, and we found that the reported trust ratings of the EMP group were significantly higher than both the single-pitch and no audio groups for the robotic arm. We found that our audio system made no significant difference in any metric when used on a humanoid robot implying audio needs to be separately designed for each platform.

Keywords: trust, audio, prosody, emotion, robotics

*** Corresponding author: Richard Savery,** Georgia Institute of Technology, Georgia Tech Center for Music Technology, Atlanta, United States, e-mail: rsavery3@gatech.edu
Lisa Zahray, Gil Weinberg: Georgia Institute of Technology, Georgia Tech Center for Music Technology, Atlanta, United States

1 Introduction

Industrial co-robotic arms are showing a significant expansion in use, which is expected to continue and grow into the foreseeable future [1]. While the use of such robotic arms expands, they still lack a standard form of non-verbal communication [2]. Many non-verbal methods to establish communication between robot arms and humans, such as haptics [3] or mixed reality [4], are costly to implement from a technical and financial perspective, requiring custom equipment and training. More recent research has shown the importance of social and emotion communication for robots [5]. For collaborative robots, displaying emotion has been shown to increase key metrics, such as the likelihood of humans to follow social norms [6], supporting better engagement with disability [7] and improving the perception of the robot as an equal human collaborator [8].

In this article, we propose that emotional musical prosody (EMP) – emotion-tagged, non-verbal audio phrases based on musical melodies – can enhance social interaction and engagement with human collaborators without requiring a change in core functionality. The ability of non-linguistic sound to display information for robotic platforms beyond trivial indicators is often underutilized, despite the use of intentional sound to deliver information in almost every device we encounter day to day [9]. Speech, however, is very commonly used in audio-based communication in robotics [10]. Here we propose that the expression of EMP by robots can have a few advantages over speech. First, linguistic speech adds a significant cognitive load to systems [11], which is often not required in human–robot interactions. Additionally, linguistic streams can easily become unintelligible [12] and detract from the interaction. And lastly, many voice systems aim to sound as human-like as possible, however, studies have shown that effort to sound like a human can negatively affect the expectations set by the end user and can decrease usage [13].

It has also been shown that displaying emotion is key for creating believable agents that humans enjoy collaborating

with ref. [14] and that prosody is effective in displaying emotions for humans and robots [10]. We propose that robotic arms are especially well positioned to improve interactions through emotion, as they are non-anthropomorphic yet are also required to act as a direct collaborator with humans. Affective non-verbal behaviour has been shown to affect HRI metrics like humans' emotional state, self-disclosure, and perceived animacy of the robot [15]. But while gestures have often been studied [16], non-linguistic forms of audio feedback are under-explored [17]. Prosody has the potential to allow the robot to communicate in a manner relatable to that of humans, but different enough from human speech to avoid the uncanny valley [17]. EMP is therefore uniquely positioned to enable better robotic communication and collaboration, capturing the advantages of sonic interaction, emotion conveyance, and avoiding the uncanny valley.

In this article, we describe our approach to generating EMP using a custom dataset of musical phrases for robotic arm interaction with humans. We evaluate these interactions firstly to confirm that there is no impact through potential distraction in collaboration with a robotic arm. We then compare EMP to single-pitch audio and no audio conditions for establishing trust, trust recovery, likeability, and the perception of intelligence and safety. Finally, we analyse the same EMP on a humanoid robot to understand if, and how, our generative system can be transferred across systems and generalized.

2 Background

2.1 Robotic arm forms of communication

The research in this article primarily focuses on how changing methods of communication can improve the perception of and interaction with robots. There is currently no standardized set of approaches for allowing robotic arms to communicate and signal their intent [18]. Gesture and gaze, which are commonly used for communication in social robotics, are not readily available to robotic arms [19]. Additionally, when these forms of communication are added to arms they require significant expense, such as extra visual displays [20], and can challenge and reducing the core functionality of the arm in the case of adding extra gestures. In robotic research, forms of non-verbal communication can generally be split into six categories: kinesics, proxemics, haptics, chronemics, vocalics, and presentation [2]. Kinesics includes communication through body movement, such as gestures [21] or facial expressions, while proxemics focuses on the robotic positioning in space, such as the distance from a human collaborator [22]. Haptics refers to touch-based methods [23], while chronemics includes subtle traits such as hesitation [24]. Presentation includes the way the robot appears, such as changes based on different behaviour [25]. The category, vocalics, includes methods such as prosody [10]. While varying movement to show intent has led to successful results [26], changes to path planning and movement dynamics are often not feasible. Another effective method for robotic arms to display their intent is through projection of the robot's future trajectory [27], however, this requires a significant investment and potential distraction to the user.

2.2 Robotics and emotion

The communication of emotion in robotics has seen a significant rise in robotics over the last 30 years, primarily in social robotics but also for robotic arms [28]. Emotion can be categorized in different ways, such as a discrete categorical manner (happiness, sadness, fear, etc.) [29], as well as continuous dimensions such as valence and arousal [30]. Emotion in robot platforms can take the role of either an input, output, or be used for internal system processing [31]. In this article, we focus on the role of emotion display or emotion as a system output and how an arm can communicate an emotion. In social robotics this has been widely researched, with outputs including facial expression [32] or head and arm movement [33]. Emotion has been more commonly used as an input to robotic arms, such as facial emotion recognition to control and change the behaviour of robotic arms [34–36]. Likewise, galvanic skin response emotional evaluation on humans has been used to impact a robot's control pattern [37]. Nevertheless, robotic arm displays of emotion beyond showing intent are widely overlooked in robotics literature.

2.3 Communication for trust and trust recovery

The display of emotion is critical for trust and increases the willingness for collaboration [38]. The lack of human trust in robotic arms can lead to underutilized interaction [39]. Trust is largely developed in the first phase of a relationship between humans and robots [40,41]. First impressions from audio and visual stimuli can also damage the ability to develop trust later on in the interaction [42]. In this work, we focus on affective trust, which is developed through emotional bonds and personal relationship, as opposed to cognitive trust which is more task focused [43].

Importantly, relationships based on affective trust are also more resilient to mistakes by either party [44].

3 Motivation

We propose the use of EMP to establish trust for robotic arms, as it has been shown as a powerful medium to convey emotions [45]. While some recent efforts to generate and manipulate robotic emotions through prosody focused on linguistic robotic communication [46], we focus on music as a driving force for generating EMP [47,48]. EMP has key use cases in situations such as where full language is not needed, or in group situations where multiple speaking voices can create linguistic confusion.

EMP as a feedback mechanism for human–robot interactions avoids the uncanny valley [17] and has the potential to allow the robot to communicate in a manner relatable to that of humans, but still different from day to day human speech. While robotic arms themselves do not generally approach the uncanny valley, we believe that independent modalities (such as a human voice) can cause the same impact on an interaction. This extends the notion of the habitability gap [49], where issues with interaction occur when a robot's functionality does not match its capability. In a robotic arm this could occur when a simple task, such as repeatedly moving an object, is accompanied by rich language-based communication method. Additionally, affective non-verbal behaviour has been shown to alter humans' emotional state and self-disclosure when interacting with a robot, as well as their perceived animacy of the robot [15]. Sonification (turning data into sound) and non-speech audio feedback have been used to improve user performance in a variety of tasks [50], implying that non-linguistic audio has a vastly underutilized potential for robotic applications.

4 Method

We conducted two different studies, one using a robotic arm and the other using a humanoid robot, in an effort to address the following research questions.

4.1 Research questions and hypotheses

Our first research question focuses on understanding the role of EMP and trust in robotic interaction.

RQ1: *How does EMP alter trust and trust recovery from mistakes, compared to no audio and single-pitch audio?*

For this question our hypothesis is that the overall trust at the end of the interaction will be significantly higher for EMP over single-pitch and higher for single-pitch audio over no audio. Our second research question compares common HRI metrics such as the perceived intelligence, perceived safety, and likeability, for each robotic system.

RQ2: *How does EMP alter perceived safety, perceived intelligence and likeability?*

For the first two research questions, we believe that participants will develop an internal model of the robot as an interactive emotional collaborator for the EMP model. This will lead to higher levels of trust and improved perception of safety and intelligence. The third question explores the relation between users' self-reported metrics, gathered through surveys and their actual responses collected through a performance-based task. We are interested in comparing whether the system that is self-reported with higher trust ratings is actually utilized more in performance-based tasks.

RQ3: *When a user indirectly self-reports higher levels of trust in a robot, does this in turn lead to higher utilization and trust in a robotic arm's suggestions?*

For these questions we hypothesize that users' self-reported trust ratings will correspond to their actual use and trust levels, as implied by choice to follow the decisions of the robotic system. We also hypothesize that by utilizing EMP, human collaborators will be more likely to trust the robotic arm's suggestions directly after a mistake.

4.2 Measures

We chose to use two existing measures for our study; Schaefer's survey for robotic trust [51], and the Godspeed measurement for Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and the Perceived Safety of Robots [52]. We used the complete Godspeed survey as described in the original article, with 24 questions to cover each metric, and each metric using a bipolar 5 point scale. For Schaefer's survey we used the 14-point version, with each question involving the participant ranking the question between 0 and 100%, which then are averaged to give a complete trust percentage.

While the Godspeed metrics have recently shown to be problematic, we chose to use them for several reasons.

One issue is the use of a bipolar-scale rating instead of a Likert scale [53]. We believe that having a high rating for Cronbach's alpha somewhat alleviates the concern that each rating is not truly opposite, as that implies that at least each rating is internally reliable. In comparison to other alternate metrics that build on the Godspeed metrics, such as the Robotic Social Attributes Scale [54], the Godspeed survey was chosen as it allows easy comparison between both our own existing studies and many previous studies that address the qualities from Godspeed [55].

4.3 Experimental design

Our experiment requires participants to perform a pattern learning and prediction task collaboratively with a robot. This is followed by the Godspeed metrics and Schaefer's survey. The study process followed the following five steps for each participant:

(1) Consent form and introduction to online form;
(2) Description of the pattern recognition task;
(3) 20 Trial Pattern Recognition Tasks;
(4) 80 Pattern Recognition Tasks, recorded for data;
(5) Godspeed and Schaefer Trust Survey (order randomized per participant);
(6) General comments and demographic information.

The pattern learning method was originally created by Van Dongen and Van Maanen to understand the reliance on decision agents and develop a framework for testing different agents [56]. Since then it has been re-purposed many times, including for comparing the dichotomy of human–human and human–automation trust [57], as well as the use of audio by cognitive agents [58].

After collecting the consent form, participants went through a description of the task, followed by 20 trial questions to teach them the process. This was followed by the recorded and analysed 80 questions. We finally allowed participants to add any general comments about the study or robot.

We modified the original pattern recognition task, asking participants to correctly predict the next number in a sequence advised by an animated model of a robot on a computer screen. Participants were told beforehand that humans and the pattern recognition software tend to be about 70% accurate on average, which has been shown to cause humans to alternate between relying on themselves and a decision agent. No further information was provided to the participants about the sequence's structure. The sequence was made up of a repeated sub-sequence that was five numbers long, containing only 1, 2, or 3 (such as 3, 1, 1, 2, 3). To prevent participants from quickly identify the pattern, 10% of the numbers in the sequence were randomly altered. Participants first completed a training exercise to learn the interface, in which a sub-sequence was repeated four times (20 total numbers). Then participants were informed a new sequence had been generated for the final task. This was generated in the same way, using a new sub-sequence with 16 repetitions (80 total numbers). Before the user chose which number they believed came next in the sequence, the robot would suggest an answer, with the robot being correct 70% of the time. This process mirrors the process from the original paper [56].

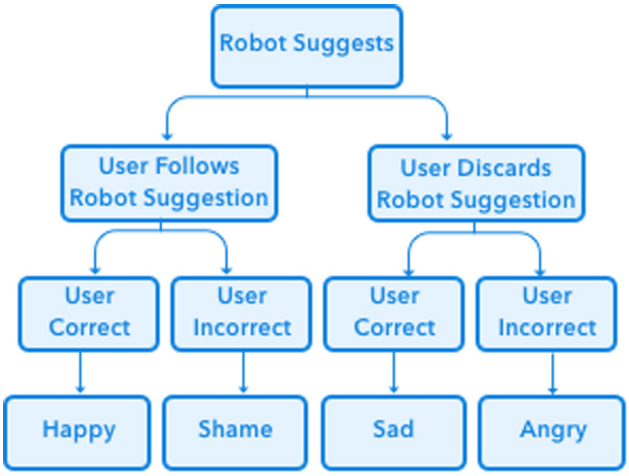


Figure 1: Robot Arm Emotional Response (bottom row indicates robot response).

The previous timestep's correct answer was displayed for the user at decision time to help them better keep track of the pattern during the animated robotic movements. We required participants to submit their answer after the robot finished pointing to its prediction, which took between 2.5 and 4.5 s. This also forced participants to spend time considering their decision given the robot's recommendation. The robot would respond to the user's choice depending on the outcome and the version of the experiment. The emotional response to a user's action with the emotion was determined by the process shown in Figure 1.

4.4 Experimental groups and robot reactions

This study was designed as a between-group experiment, where participants were randomly allocated to one of the

three groups. These groups were an EMP audio group, a single-pitch audio group (notes), and a control with no audio (gesture). In all three versions of the experiment, the robot responded with the emotional gestures described in Section 4.8.

In the EMP group, the gestural response was accompanied by playing a prosody-based audio sample, randomly selected each time from the five phrases matching the response emotion. These phrases were obtained using the process described in Section 4.6. In the notes group, the gestural response was accompanied by playing one musical note. Each emotion was randomly assigned one pitch from the midi pitches 62, 65, 69, and 72. The notes were chosen as they are in both the male and female vocal range and a similar pitch range to the EMP. We chose not to alternate timbre (the audio features outside pitch) for this group, as the EMP group already contained significant timbre variety. This assignment remained consistent throughout the experiment to maintain a relation between the sounds and the outcome. For each pitch, five different audio files were available to be selected, each with a different instrument timbre and length (varying from 2 to 5 s), to provide variety similar to that of the five different EMP phrases available for each emotion. Finally, in the gesture group, the gesture was performed in silence.

4.5 Participants

For each of the studies, we recruited 46 participants through the online survey platform Prolific for a total of 92 participants. The participants, ages ranged from 19 to 49, while the mean age was 25, with a standard deviation of 7. Participants were randomly sorted into one of the categories – audio with EMP (15 participants), single-pitch audio (16 participants), and no audio (15 participants). Each experiment took approximately 30 min to complete. Participants were paid $4.75 USD.

4.6 Dataset

In previous study, we created a deep learning generative system for EMP [47,48]. For this article and experiment, we chose to use our recently created dataset of a human singing emotional phrases, to avoid any potential noise added by a generative system. The recorded dataset contains 4.22 h of musical material recorded by Mary Carter, divided into 1–15 s phrases each corresponding to one of the 20 different emotions in the Geneva Emotion Wheel [59] shown in Figure 2.

Forty-five participants from Prolific and Amazon Mechanical Turk (MTurk) validated our generative system,

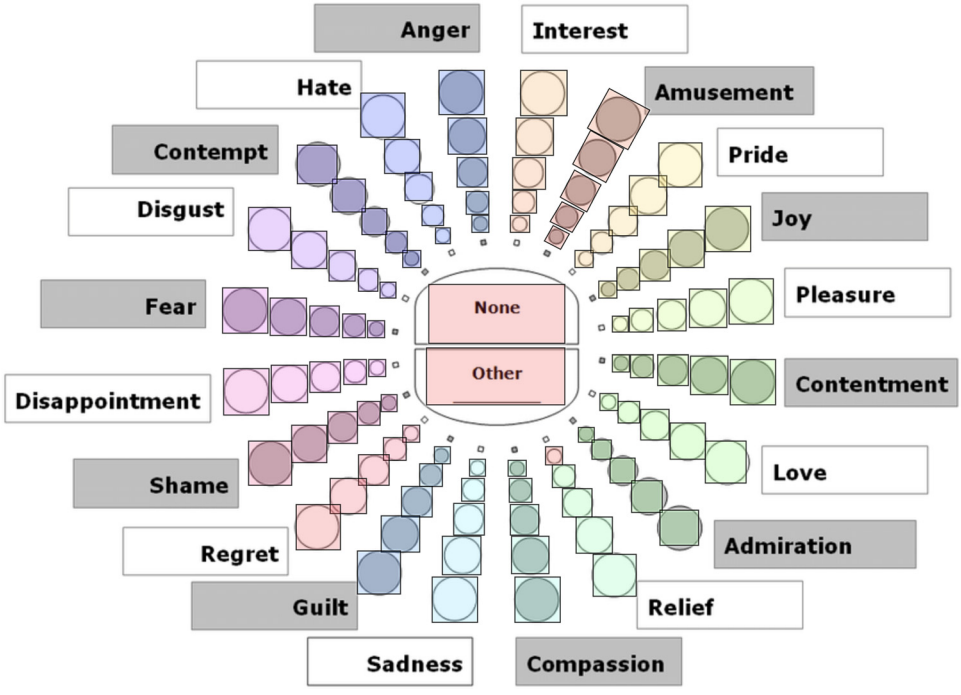


Figure 2: Geneva emotion wheel.

by selecting an emotion and intensity when listening to each provided phrase. For quality assurance, participants were randomly given test questions throughout the experiment asking them to select a certain answer. Each participant was given 6.5 on average, and responses which had more than one incorrect attention question were ignored, leaving a total of 45 participants for data analysis. In order to minimize the length of the survey, questions were randomly allocated, with 12 participants on average evaluating each individual phrase. Answers of "None" or "Other" were ignored in the analysis, resulting in an average of 11.3 valid evaluations for each phrase.

Our analysis of the phrases used the metrics defined by Coyne et al. [60]. We calculated the rated emotion's mean and variance in units of emotion (converted from degrees on the wheel), weighted by user-rated intensity.

In the experiment, we used phrases for the four emotions: joy, shame, sadness, and anger. These emotions were chosen to best match the outcomes in Figure 1 using gesture descriptions specified in ref. [61]. Five phrases for each emotion were chosen to add variety to the robot's response in an effort to prevent tiring the user with the same sounds, and to control for any preference for individual phrases. In selecting the phrases for each of the four emotions, phrases from the closest two other emotions on the wheel within the same quadrant were also considered for selection. The sets were {joy, pride, pleasure}, {shame, disappointment, regret}, {sadness, guilt, regret}, and {anger, hate, contempt}. We selected 5 of the 15 potential phrases for each trial by limiting length to be between 4 and 10 s. This restricted the variance to be less than 2, requiring the weighted mean emotion rating to fall within the correct quadrant of the wheel. The selected phrases were the ones with the smallest difference between the actual emotion and mean-rated emotion.

4.7 Interaction

Participants interacted with a virtual 3D model of the robot in an application designed in Unity. This allowed us to have interactions and responses vary based on user choices, while leveraging the quantity of participants available for an online study. Each time a participant was asked to answer a question, the robot acted as a decision agent, pointing to an answer that may be correct or incorrect. The user would then type their answer using their computer keyboard. We used three versions of interaction, varying the way the robot reacted to the user's answer, as described in Section 4.4. An example image of the robotic arm interface is shown in Figure 3.

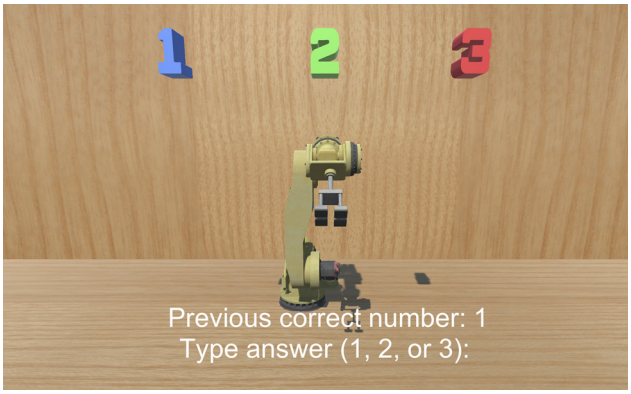


Figure 3: Example image from the robot interaction application.

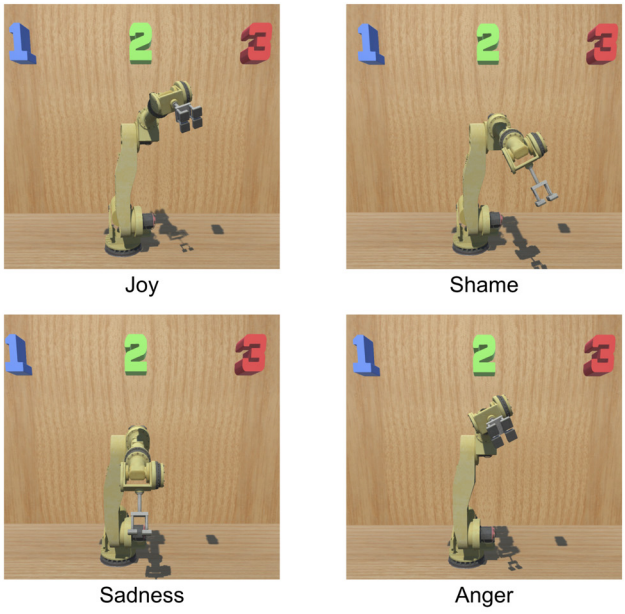
4.8 Gestures

We created a gesture for each of the emotions joy, shame, sadness, and anger. The gestures were designed according to the table of emotion-specific non-verbal behaviours provided in ref. [61] as well as our own *post hoc* overview of discriminative body movements and poses. This approach has been used before in designing emotional robot gestures [62]. For the humanoid embodiment, we were able to incorporate more specific body language such as forming hands into fists and simulating crying.

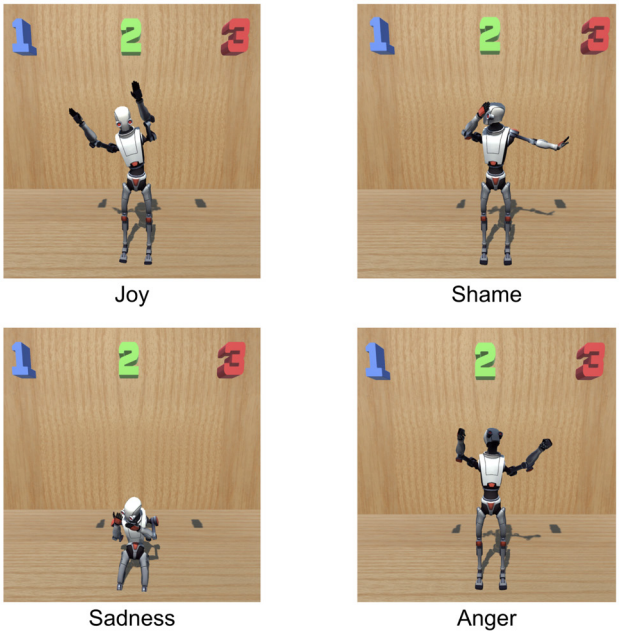
Our Joy gesture has the arm lift up high, making three quick upwards movements alternating which side it faces. The humanoid lifts both of its arms up and waves them back and forth, repeats this motion with its arms higher, and finally jumps into the air. For Shame the arm slowly bends down and away from the camera to one side, while the humanoid looks to one side and moves its hand to cover its face. For Sadness, the arm slowly bends down while still centred with respect to the camera, while the humanoid falls to its knees and covers its face with both hands. The Anger gesture has the arm first lean downwards and make two fast lateral movements, and then lean upwards to make two more fast lateral movements. The humanoid raises its fists into the air and push its torso forward. Examples of poses encountered during each gesture are shown in Figure 4.

5 Results

To answer our research questions we used metrics from the trust survey, Godspeed measure, and the amount of times participants accepted the robot's suggestion. Research question 1 analyses the results from the trust survey, while research question 2 focuses on the Godspeed metrics.



(a)



(b)

Figure 4: Example poses passed through during emotional gestures: (a) arm and (b) humanoid.

Research question 3 compares the results from the trust survey with participant choices throughout the experiment.

5.1 RQ1: Trust recovery

Research question 1 analysed how trust varies by audio type for each robotic platform. We hypothesized that trust would improve on both platforms with the use of EMP.

5.1.1 RQ1: Arm

We first calculated Cronbach's alpha for each metric in the trust survey, which gave a high reliability of 0.92. We then calculated the overall trust score by inverting categories when appropriate and then generating the mean for each individual. The mean trust of each group was EMP 71%, notes 57%, and gesture 62% (Figure 5). After running a one-way ANOVA the p-value was significant, $p = 0.041$, $f = 3.4$. Pair-wise t-tests between groups' trust rating gave the results: notes-gestures $p = 0.46$, notes-EMP $p = 0.025$, and gesture-EMP $p = 0.025$. This supports our hypothesis that trust would be higher from the arm using EMP.

We also evaluated trust based on participants' actual use of the system. The percentage of answers for which users agreed with the robot for each group is plotted in Figure 6a. We performed a one-way ANOVA test to test whether there was a significant difference in this metric between groups, $p = 0.68$, which was not significant.

To compare trust recovery after mistakes between groups, we analysed the percentage of times each user agreed with the robot immediately after an instance of following the robot's incorrect suggestion. The results are plotted in Figure 6a. The one-way ANOVA test yielded $p = 0.87$, which was not significant.

5.1.2 RQ1: Humanoid

Cronbach's alpha for the humanoid trust survey was 0.89, showing a high internal consistency. We followed the same procedure to calculate the trust scores, with the means 63% for notes, 64% for gesture, and 66% for EMP (Figure 5). Running a one-way ANOVA and pair-wise t-tests showed no significance ($p > 0.05$).

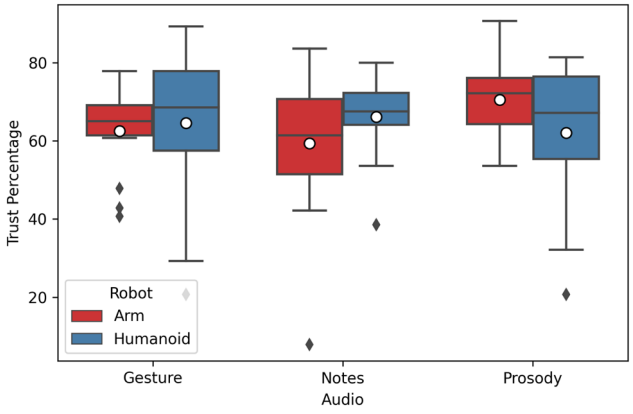


Figure 5: Box plot of Trust metrics. White dot indicates mean, middle line is median, and black dots are outliers.

Figure 6b shows the results for percent agreement with the robot, and percent agreement with the robot after it made a mistake. A one-way ANOVA between groups for percentage of answers in which users agreed with the robot yielded $p = 0.0039$, which was significant. A two-tailed t-test between each pair of groups had significant results for gestures versus EMP at $p = 0.0021$ and gestures versus notes at $p = 0.018$. The one-way ANOVA for percent agreement after the robot's mistake was not significant, with $p = 0.13$. We note that we did not remove outliers for these statistical tests due to the number of participants in each group.

5.2 RQ2: Anthropomorphism, safety, intelligence, and likeability

Research question 2 identified how EMP, notes, or gesture alone varied each Godspeed metric for the arm and humanoid robot. Cronbach's alpha for the robotic arm result in Anthropomorphism (0.85), Intelligence (0.89), and Likeability (0.92), and all showed high reliability values above 0.85. Safety's coefficient was slightly lower at 0.75. For the humanoid calculating Cronbach's alpha for anthropomorphism, intelligence, and likeability gave 0.80, 0.90, and 0.88, respectively, demonstrating high reliability. Safety's Cronbach alpha however resulted in 0.50, indicating the survey did not present internal validity. Due to the low internal reliability we chose not to analyse the safety results. This is discussed further in Section 6.4.

5.2.1 RQ2: Arm

We first performed a one-way ANOVA for each category, which showed no significant results. Performing paired t-tests with Holm–Bonferroni corrections showed significance for anthropomorphism between EMP and gesture ($p = 0.048$) and EMP and notes ($p = 0.003$). Likeability was also significant between notes and EMP ($p = 0.048$). Figures 7–9 show box plots for anthropomorphism, intelligence, and likeability. This did not support our hypothesis as we were unable to show difference between audio types for safety or likeability across all categories.

5.2.2 RQ2: Humanoid

Across each audio category the humanoid achieved very similar results between the audio and gesture variables, with no significant difference. For example, likeability received ratings of 3.5, 3.52, and 3.61 for notes, gesture, and EMP. These results indicated that the audio used made no difference to the perception of the robot. Figures 7–9 show the results for each metric.

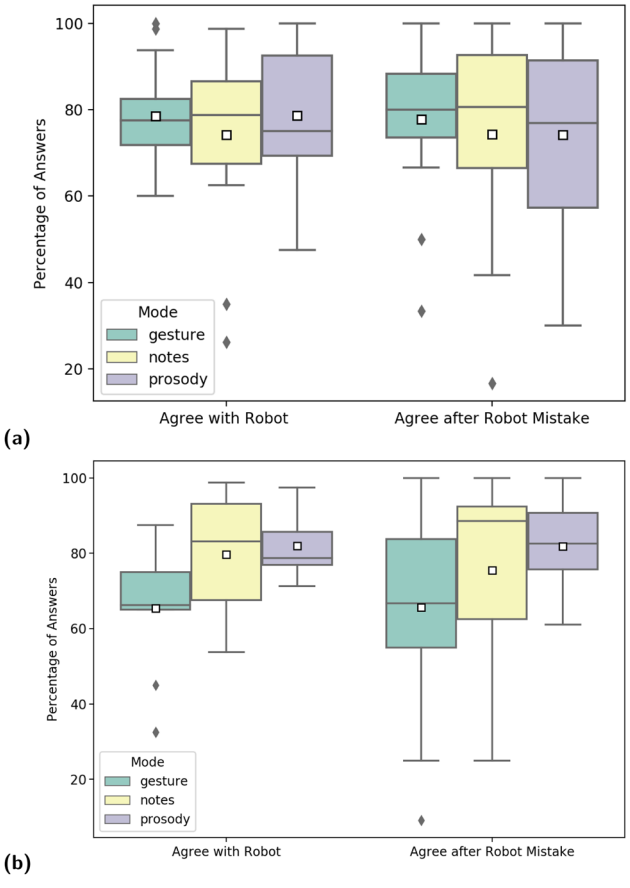


Figure 6: Box plots showing percentage of answers agreeing with the robot overall and after the robot made a mistake (means indicated by white squares): (a) arm, (b) humanoid.

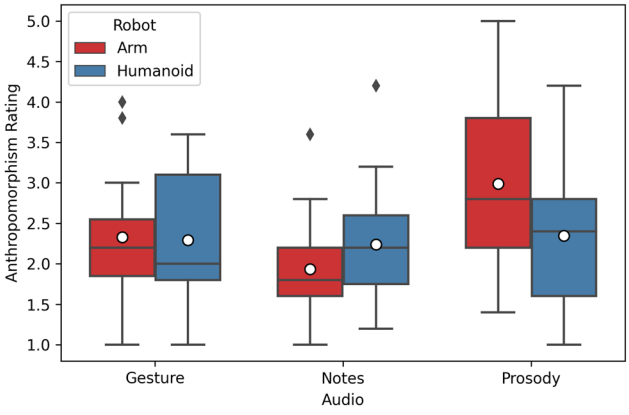


Figure 7: Box plot of anthropomorphism, comparing humanoid and arm across audio types. White dot indicates mean, middle line is median, and black dots are outliers.

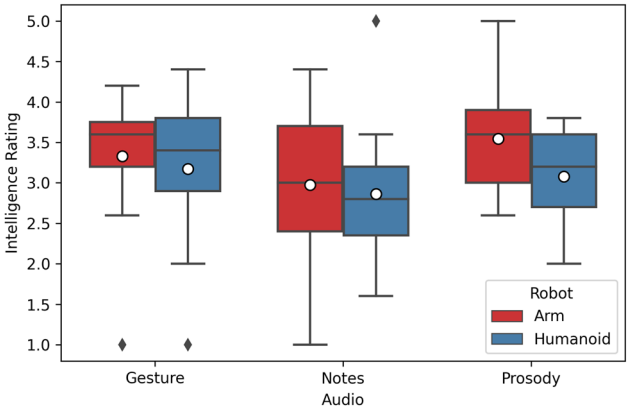


Figure 8: Box plot of perceived Intelligence, comparing humanoid and arm across audio types. White dot indicates mean, middle line is median, and black dots are outliers.

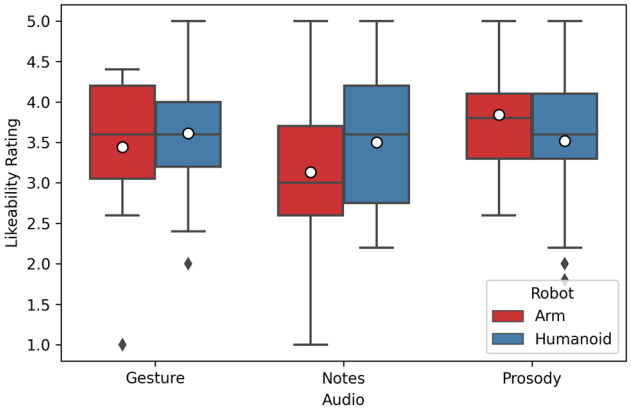


Figure 9: Box plot of likeability, comparing humanoid and arm across audio types. White dot indicates mean, middle line is median, and black dots are outliers.

5.3 RQ3: Trust survey and participant choices

Research question 3 explored the relationship between the trust survey and participants, actual choices throughout the experiment. We calculated the Pearson correlation coefficient between the final trust scores for the robotic arm, and the percentage of answers users agreed with the robot. The result was $r = 0.12$, which indicates a weak correlation between the two metrics.

5.4 User comments

5.4.1 Arm

The free input textual comments provided by participants indicate that it was possible, in all groups, to perceive the emotions the robot was trying to convey. In the EMP group, one user said, "The arm seems quite emotional! When it's right it is quite happy, but when it is wrong it gets particularly sad." In the notes group, a user said "When we got the right answer the robot seemed cheerful, as opposed to when we selected the wrong answer (based on the robot's recommendation) it seemed as if he was sorry for giving wrong suggestions. If I chose an option different than the robot's suggestion and its answer was correct, it seemed as if he gave the look of I told you the right answer!" And in the gesture group, one comment was "the emotions were very easily perceivable." Two participants in the notes group had negative comments on the audio response, describing it as "horrible" and "annoying," while one participant in the EMP group said the "humming was annoying." Several participants mentioned that the robot moved too slowly. Some comments mentioned having a hard time detecting any pattern in the sequence, while in others, users discussed their strategies.

5.4.2 Humanoid

In the EMP group, one user said, "It was clearly a robot (the cartoon), but the audio queues made it seem more humanlike," with another user describing the robot as "friendly." However, another user in this group described the robot as "irritating," and another explained that it was pleasant at first but became annoying over time. In the notes group, two users used the phrase "over the top" when describing the robot's reactions. Two other users mentioned that the robot seemed excited or like it was having a good time. One user said "I feel like the sound effects aren't really necessary." In the gestures group, one user said "the robot seemed really happy when i got things right, but when i kept failing consistently i felt i was embarrassing it/letting it down, which added more pressure to me to get it [sic] right." Two other users described a similar interpretation of the reactions. Two different users in this group mentioned that the robot seemed rigid or mechanic. Users' discussions of their strategies varied from trusting the robot most of the time to trusting their own instincts more than the robot.

5.5 Summary

Our results indicated significant results for multiple areas for the robotic arm, however, showed no significance for

the humanoid robot. Considering trust we found EMP had significant results over the notes condition and gesture alone, when considering the responses from the user survey. For research question 2 we found significant results only for anthropomorphism on the arm between audio types.

6 Discussion

Building from our findings that EMP was able to improve trust for the robotic arm but not the humanoid, we developed multiple discussion points. Firstly, we believe the contrasting results between the humanoid and robotic arm points to the need for further research into specific audio design aimed at individual platforms. Secondly, despite the improvement in trust, we did not find variations in the likeability rating of the robot. Finally, the link between our survey response for trust and the users choice to follow a robots suggestion was not clear and requires further investigation into what is being measured by the trust scale.

6.1 Platform-specific audio design

Our goal for embedding EMP in robots was to develop and evaluate a non-language-based form of audio communication, which could help avoid the uncanny valley. While we were successful in improving trust through embedding EMP in robotic arms, in humanoid robots we found no significant improvement in any category. This could be generally interpreted as meaning that audio does not alter humanoid perception as much as for a lower degree of freedom, non-anthropomorphic, robotic arms. It can also be claimed that our particular audio synthesis implementation did not lead to the desired results in humanoid robots, but that other future implementations might. The category humanoid robot is also very broad, with potential that the humanoid model we used was not able to be modified with audio, or that a feature such as the eyes dominated user perception.

The most surprising result was that the pitch audio fell well below the median of gestures-only in every category. This may indicate that while EMP can lead to positive outcomes, audio when implemented ineffectively has the capability of drastically reducing HRI metrics. The reason for this is likely due to the fact that the notes' sound was not related to the emotion being displayed by the gesture beyond remaining consistent throughout the experiment.

In any case, these results reiterated that audio must be carefully considered for every platform, without only reusing existing speech systems. More broadly, we believe this work indicates the importance of audio design in robotics, and the impact that robotic audio can have on human perception. Through changing audio alone and not relying on default audio methods such as speech, we were able to drastically change the perception of a robotic system, but have no impact on a different system. While we have shown EMP as particularly effective at improving a range of metrics, this is just one of the many possible approaches that could be developed within more careful future audio design for human–robot interaction.

6.2 Anthropomorphism, likeability, and trust

Comparing the Godspeed metrics, it was unsurprising to find that the addition of human vocalizations increased the anthropomorphism of the arm. We had expected likeability to become higher, and while it was not a significant result, it would still be worth investigating further with more subjects. The relationship between each of these metrics is complex, with no clear relation between likeability and trust, or how EMP alters a human collaborators, perspective on these metrics. In our past research we have also found a complicated relationship between which features can be improved by EMP, such as an improved rating of the functionality of a robot in creative settings [63], but not functionality in industrial tasks [64].

6.3 Measuring trust

Users' ratings of trust in the survey did not strongly correlate with their actual behaviour during the task in terms of how often they agreed with the robot's suggestions. This is consistent with the fact that while users reported significantly higher trust for audio with musical prosody, no significant differences were found in their actual choices during the interactions. A similar conflict between these types of metrics was found in the original decision framework article [56], where higher reported trust in the arm did not always result in higher percent agreement with the arm.

We believe the primary reason for the contrasting rating for trust and how often participants agreed with

the robot, is due to the multifaceted nature of trust itself. We contend that EMP is most impactful for changing ratings for affective trust, a type of trust that develops through emotion and social relationships. This contrasts with cognitive trust, which is based on a user's actual willingness to trust or rely on a collaborator to perform a task [65]. In robotics, trust has similarly been broken into performance trust and moral trust [66]. Performance trust occurs where a human collaborator believes the system is capable of performing the required action. The counter, moral trust, is a rating of the collaborators belief the robot desires to perform the morally correct task. While we make no claim that either measure we utilized to gauge trust directly correlates with a type of trust, we believe the trust survey is more likely to lead towards high ratings for affective or moral trust.

6.4 Measuring perceived safety

While the Godspeed survey has been extensively used in HRI with 1,306 citations by December 2020, we believe an online study with animations may not have effectively capture the metrics used for perceived safety. We found participants often described themselves as calm on the Godspeed scale, but also surprised, likely due to the online setting where surprising gestures did not change a participant's self-perception as calm.

6.5 Limitations

This study was performed using virtual interactions with a robot and 46 participants. It would be useful to investigate this further with a larger sample size, and to have participants interact with a physical robot for comparison. Additionally, more variations of robot responses could be compared and analysed beyond the three that we investigated. For example, prosodic audio of a human voice could be compared with that of musical instruments.

7 Conclusion

This research demonstrates that when the robot arm model responded with EMP users reported higher trust metrics than when the robot responded with single-pitched notes or no audio. This supports our hypothesis that EMP has a positive effect on humans' trust of a robotic arm. Our additional findings and discussion points support the complex relationship between trust metrics and how users interact with robots, as well as the challenges in measuring trust itself. We did not find significant results for likeability, anthropomorphism, or perceived intelligence through prosody, although the arm with prosody did achieve higher means across both categories. In studies with a humanoid robot we found no significant changes in metrics, with audio seemingly have no impact on ratings. This indicates that audio design is a crucial step for human–robot interaction and can not simply be transferred between platforms without consideration of the broader impact.

Acknowledgements: This material is based upon work supported by the National Science Foundation under Grant No. 1925178. The authors would like to thank Destiny Wellington-Wilson and Varshita Patakottu who designed the gestures used for the humanoid robot.

Funding information: This material was based upon work supported by the National Science Foundation under Grant No. 1925178.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: Authors state no conflict of interest.

Informed consent: Informed consent was obtained from all individuals included in this study.

Ethical approval: The research related to human use has been complied with all the relevant national regulations, institutional policies and in accordance the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board.

Data availability statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

[1] Grand View Research Choice, "Collaborative robots market size, share and trends analysis report by payload capacity, by application (assembly, handling, packaging, quality testing),

by vertical, by region, and segment forecasts, 2019–2025," Grand View Research Choice, Technical Report, 2018.
[2] S. Saunderson and G. Nejat, "How robots influence humans: A survey of non-verbal communication in social human–robot interaction," *Int. J. Soc. Robot.*, vol. 11, no. 4, pp. 575–608, 2019.
[3] M. Tannous, M. Miraglia, F. Inglese, L. Giorgini, F. Ricciardi, R. Pelliccia, et al., "Haptic-based touch detection for collaborative robots in welding applications," *Robot. Comput. Integr. Manuf.*, vol. 64, art. 101952, 2020.
[4] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, G. Konidaris, et al., "Communicating robot arm motion intent through mixed reality head-mounted displays," in *Robotics Research*, Providence, RI, USA: Springer, 2020, pp. 301–316.
[5] K. Fischer, "Why collaborative robots must be social (and even emotional) actors," *Tech. Res. Philos. Technol.*, vol. 23, no. 3, pp. 270–289, 2019.
[6] J. Jost, T. Kirks, S. Chapman, and G. Rinkenauer, "Examining the effects of height, velocity and emotional representation of a social transport robot and human factors in human–robot collaboration," in *IFIP Conference on Human-Computer Interaction*, Paphos, Cyprus: Springer, 2019, pp. 517–526.
[7] S. S. Balasuriya, L. Sitbon, M. Brereton, and S. Koplick, "How can social robots spark collaboration and engagement among people with intellectual disability?," in *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, ser. OZCHI'19, New York, NY, USA: Association for Computing Machinery, 2019, pp. 209–220, DOI: https://doi.org/10.1145/3369457.3370915.
[8] L. Desideri, C. Ottaviani, M. Malavasi, R. di Marzio, and P. Bonifacci, "Emotional processes in human–robot interaction during brief cognitive testing," *Comput. Human Behav.*, vol. 90, pp. 331–342, 2019.
[9] B. N. Walker and G. Kramer, "Human factors and the acoustic ecology: Considerations for multimedia audio design," in *Audio Engineering Society Convention 101*, Los Angeles, California: Audio Engineering Society, 1996.
[10] J. Crumpton and C. L. Bethel, "A survey of using vocal prosody to convey emotion in robot speech," *Int. J. Soc. Robotic.*, vol. 8, no. 2, pp. 271–285, 2016.
[11] J. Lopes, K. Lohan, and H. Hastie, "Symptoms of cognitive load in interactions with a dialogue system," in *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*, 2018, pp. 1–5.
[12] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acustica United with Acustica*, vol. 86, no. 1, pp. 117–128, 2000.
[13] B. R. Cowan, N. Pantidi, D. Coyle, K. Morrissey, P. Clarke, S. Al-Shehri, et al., "'What can i help you with?': Infrequent users' experiences of intelligent personal assistants," in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI '17, New York, NY, USA: Association for Computing Machinery, 2017, DOI: https://doi.org/10.1145/3098279.3098539.
[14] M. Mateas, "An Oz-centric review of interactive drama and believable agents," in *Artificial Intelligence Today: Recent Trends and Developments*, M. J. Wooldridge and M. Veloso, Eds., Berlin, Heidelberg: Springer-Verlag, 1999, pp. 297–328. http://dl.acm.org/citation.cfm?id=1805750.1805762
[15] A. M. Rosenthal-vonder Pütten, N. C. Krämer, and J. Herrmann, "The effects of humanlike and robot-specific affective nonverbal behaviour on perception, emotion, and behaviour," *Int. J. Soc. Robot.*, vol. 10, no. 5, pp. 569–582, 2018.
[16] A. Beck, L. Cañamero, and K. A. Bard, "Towards an affect space for robots to display emotional body language," in *Roman 2010*, IEEE, 2010, pp. 464–469.
[17] K. F. MacDorman, "Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley," in *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, 2006, pp. 26–29.
[18] E. Cha, Y. Kim, T. Fong, and M. J. Mataric, "A survey of nonverbal signaling methods for non-humanoid robots," *Foundat. Trends® Robot.*, vol. 6, no. 4, pp. 211–323, 2018.
[19] E. Rosen, D. Whitney, E. Phillips, G. Chien, J. Tompkin, et al., "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," *Int. J. Robot. Res.*, vol. 38, no. 12–13, pp. 1513–1526, 2019.
[20] W. Jitviriya and E. Hayashi, "Design of emotion generation model and action selection for robots using a self organizing map," in *2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (*ECTI-CON*), 2014, pp. 1–6.
[21] B. Gleeson, K. MacLean, A. Haddadi, E. Croft, and J. Alcazar, "Gestures for industry intuitive human–robot communication from human observation," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction* (*HRI*). Tokyo, Japan: IEEE, 2013, pp. 349–356.
[22] M. L. Walters, K. Dautenhahn, R. te Boekhorst, K. L. Koay, C. Kaouri, et al., "The influence of subjects' personality traits on personal spatial zones in a human–robot interaction experiment," in *ROMAN 2005, IEEE International Workshop on Robot and Human Interactive Communication, 2005*, Nashville, TN, USA: IEEE, 2005, pp. 347–352.
[23] H. Fukuda, M. Shiomi, K. Nakagawa, and K. Ueda, "Midas touch'in human–robot interaction: Evidence from event-related potentials during the ultimatum game," in *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2012, pp. 131–132.
[24] A. Moon, C. A. Parker, E. A. Croft, and H. M. Van der Loos, "Did you see it hesitate? – Empirically grounded design of hesitation trajectories for collaborative robots," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, San Francisco, CA, USA: IEEE, 2011, pp. 1994–1999.
[25] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behaviour to tasks to improve human–robot cooperation," in *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003, Proceedings, ROMAN 2003*, Millbrae, CA, USA: IEEE, 2003, pp. 55–60.
[26] C. Bodden, D. Rakita, B. Mutlu, and M. Gleicher, "Evaluating intent-expressive robot arm motion," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication* (*RO-MAN*), New York, NY, USA: IEEE, 2016, pp. 658–663.
[27] E. Ruffaldi, F. Brizzi, F. Tecchia, and S. Bacinelli, "Third point of view augmented reality for robot intentions visualization," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, Lecce, Italy: Springer, 2016, pp. 471–478.

[28] R. Savery and G. Weinberg, "Robots and emotion: a survey of trends, classifications, and forms of interaction," *Adv. Robot.*, vol. 35, no. 17, pp. 1030–1042, 2021, DOI: https://doi.org/10.1080/01691864.2021.1957014.
[29] L. Devillers, L. Vidrascu, and L. Lamel, "Challenges in real-life emotion annotation and machine learning based detection," *Neural Netw.*, vol. 18, no. 4, pp. 407–422, 2005.
[30] J. A. Russell, "Emotion, core affect, and psychological construction," *Cognit. Emotion*, vol. 23, no. 7, pp. 1259–1283, 2009.
[31] R. Savery and G. Weinberg, "A survey of robotics and emotion: Classifications and models of emotional interaction," in *Proceedings of the 29th International Conference on Robot and Human Interactive Communication*, 2020, pp. 986–993.
[32] D.-S. Kwon, Y. K. Kwak, J. C. Park, M. J. Chung, E.-S. Jee, et al., "Emotion interaction system for a service robot," in *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, Jeju Island, South Korea: IEEE, 2007, pp. 351–356.
[33] J. Li and M. Chignell, "Communication of emotion in social robots through simple head and arm movements," *Int. J. Soc. Robot.*, vol. 3, no. 2, pp. 125–142, 2011.
[34] Y. Mei, "Emotion-driven attention of the robotic manipulator for action selection," in *2016 35th Chinese Control Conference* (*CCC*), Chengdu, China: IEEE, 2016, pp. 7173–7178.
[35] S. Iengo, A. Origlia, M. Staffa, and A. Finzi, "Attentional and emotional regulation in human–robot interaction," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, Paris, France: IEEE, 2012, pp. 1135–1140.
[36] M. Ying and L. Zhentao, "An emotion-driven attention model for service robot," in *2016 12th World Congress on Intelligent Control and Automation* (*WCICA*), Guilin, China: IEEE, 2016, pp. 1526–1531.
[37] Y. Takahashi, N. Hasegawa, K. Takahashi, and T. Hatakeyama, "Human interface using PC display with head pointing device for eating assist robot and emotional evaluation by GSR sensor," in *Proceedings 2001 ICRA, IEEE International Conference on Robotics and Automation* (*Cat. No. 01CH37164*), vol. 4, Seoul, South Korea: IEEE, 2001, pp. 3674–3679.
[38] T. Gompei and H. Umemuro, "Factors and development of cognitive and affective trust on social robots," in S. Ge et al. (eds), *Social Robotics, ICSR 2018, Lecture Notes in Computer Science*, vol. 11357, Springer, Cham, 2018, pp. 45–54.
[39] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004.
[40] P. H. Kim, K. T. Dirks, and C. D. Cooper, "The repair of trust: A dynamic bilateral perspective and multilevel conceptualization," *Academy Manag. Rev.*, vol. 34, no. 3, pp. 401–422, 2009.
[41] R. E. Miles and W. D. Creed, "Organizational forms and managerial philosophies-a descriptive and analytical review," *Res. Org. Behav. Ann. Ser. Anal. Essay. Critic. Rev.*, vol. 17, pp. 333–372, 1995.
[42] K. E. Schaefer, *Measuring Trust in Human Robot Interactions: Development of the Trust Perception Scale-HRI*, Boston, MA: Springer US, 2016, pp. 191–218.
[43] A. Freedy, E. De Visser, G. Weltman, and N. Coeyman, "Measurement of trust in human–robot collaboration," in *CTS 2007 International Symposium on Collaborative Technologies and Systems, 2007*, Orlando, Florida: IEEE, 2007, pp. 106–114.
[44] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer, "Not so different after all: A cross-discipline view of trust," *Acad. Manag. Rev.*, vol. 23, no. 3, pp. 393–404, 1998.
[45] J. Sloboda, "Music: Where cognition and emotion meet," in *Conference Proceedings: Opening the Umbrella; an Encompassing View of Music Education; Australian Society for Music Education, XII National Conference, University of Sydney, NSW, Australia, 09–13 July 1999*, Sydney, Australia: Australian Society for Music Education, 1999, p. 175.
[46] C. Breazeal and L. Aryananda, "Recognition of affective communicative intent in robot-directed speech," *Autonom. Robot.*, vol. 12, no. 1, pp. 83–104, 2002.
[47] R. Savery, R. Rose, and G. Weinberg, "Finding Shimi's voice: fostering human–robot communication with music and a NVIDIA Jetson TX2," *Proceedings of the 17th Linux Audio Conference*, (LAC-19), CCRMA, Stanford University, USA, March 23–26, 2019.
[48] 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*), New Dehli, India: IEEE, 2019, pp. 1–7.
[49] R. K. Moore, "Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction," in *Dialogues with Social Robots*, Springer, 2017, pp. 281–291.
[50] B. N. Walker and M. A. Nees, "Theory of Sonification," in: *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff, Eds., Logos Verlag Berlin, 2011, ch. 2.
[51] K. E. Schaefer, "Measuring trust in human robot interactions: Development of the 'trust perception scale-HRI,'" in *Robust Intelligence and Trust in Autonomous Systems*, Springer, 2016, pp. 191–218.
[52] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *Int. J. Soc. Robot.*, vol. 1, no. 1, pp. 71–81, 2009.
[53] A. Kaplan, T. Sanders, and P. Hancock, "Likert or not? How using Likert rather than biposlar ratings reveal individual difference scores using the Godspeed scales," *Int. J. Soc. Robot.*, vol. 13, pp. 1553–1562, 2021.
[54] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The Robotic Social Attributes Scale (RoSAS) development and validation," in *Proceedings of the 2017 ACM/IEEE International Conference on Human–Robot Interaction*, 2017, pp. 254–262.
[55] A. Weiss and C. Bartneck, "Meta analysis of the usage of the Godspeed questionnaire series," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication* (*RO-MAN*), IEEE, 2015, pp. 381–388.
[56] K. Van Dongen and P.-P. Van Maanen, "A framework for explaining reliance on decision aids," *Int. J. Human-Comput. Stud.*, vol. 71, no. 4, pp. 410–424, 2013.
[57] E. J. de Visser, F. Krueger, P. McKnight, S. Scheid, M. Smith, et al., "The world is not enough: Trust in cognitive agents," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, no. 1, Los Angeles, CA: Sage Publications Sage CA, 2012, pp. 263–267.

[58] L. Muralidharan, E. J. de Visser, and R. Parasuraman, "The effects of pitch contour and flanging on trust in speaking cognitive agents," in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, Toronto, Canada: ACM, 2014, pp. 2167–2172.

[59] V. Sacharin, K. Schlegel, and K. Scherer, *Geneva Emotion Wheel Rating Study* (*Report*), Geneva, Switzerland: University of Geneva, Swiss Center for Affective Sciences, 2012.

[60] A. K. Coyne, A. Murtagh, and C. McGinn, "Using the Geneva Emotion Wheel to measure perceived affect in human–robot interaction," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI 20, New York, NY, USA: Association for Computing Machinery, 2020, pp. 491–498, DOI: https://doi.org/10.1145/3319502.3374834.

[61] H. G. Walbott, "Bodily expression of emotion," *Europ. J. Soc. Psychol.*, vol. 28, no. 6, pp. 879–896, 1998.

[62] M. Bretan, G. Hoffman, and G. Weinberg, "Emotionally expressive dynamic physical behaviours in robots," *Int. J. Human-Comput. Stud.*, vol. 78, pp. 1–16, 2015.

[63] R. Savery, L. Zahray, and G. Weinberg, "Before, between, and after: Enriching robot communication surrounding collaborative creative activities," *Front. Robot. AI*, vol. 8, art. 662355, 2021.

[64] R. Savery, A. Rogel, and G. Weinberg, "Emotion musical prosody for robotic groups and entitativity," in *2021 30th IEEE International Conference on Robot & Human Interactive Communication* (*RO-MAN*), IEEE, 2021, pp. 440–446.

[65] D. Johnson and K. Grayson, "Cognitive and affective trust in service relationships," *J. Business Res.*, vol. 58, no. 4, pp. 500–507, 2005.

[66] B. F. Malle and D. Ullman, "A multidimensional conception and measure of human–robot trust," in *Trust in Human-Robot Interaction*, Elsevier, 2021, pp. 3–25.