11

Augmenting a Group of Task-Driven Robotic Arms with Emotional Musical Prosody

Richard Savery, Amit Rogel and Gil Weinberg

11.1	Introduction		220
11.2	Background		221
	11.2.1	HRI and Emotion	221
	11.2.2	Sound and HRI	222
	11.2.3	Emotional Musical Prosody	222
11.3	Study 1: Fundamentals of Emotion Contagion, Music and Robotics		223
	11.3.1	Method	224
	11.3.2	Results	225
	11.3.3	Discussion	227
11.4	Study 2: Exploration Research into the Role of EMP in Groups		228
	11.4.1	Perception of Robot Emotion, Musical Reactions and Gesture	229
	11.4.2	Robotic Groups	231
	11.4.3	HRI Metrics	232
11.5	Discussion		233
	11.5.1	Memorable Moments around Musical Interaction	233
	11.5.2	Music Adds Emotion	233
	11.5.3	Embodied Music	234
	11.5.4	Music in Groups	234
	11.5.5	Music and HRI	234
	11.5.6	Limitations	234
11.6	Conclusion		235
	Bibliography		235

DOI: 10.1201/9781003320470-11

11.1 Introduction

Research at the intersection of music, emotion, and robotics has focused on work in robotic musicianship and human–robot interaction (HRI) studies. In robotic musicianship, robots are designed to perform and compose music acting as a musical collaborator [35]. Research in music and HRI however, has focused on methods for music or sound to improve how humans interact with robots [22]. In this work, we focus on how embedding musical features and gestures into robotic systems can alter the interaction and improve key HRI collaboration metrics, including trust, warmth and competence. We expand past work in the field by looking at the intersection of two rarely addressed areas, large groups of robots and the impact of sound on interaction. We contend that music, as one of the most emotionally driven human forms of communication, can play a key role in HRI.

Emotional musical prosody (EMP), where short musical phrases are used to convey an emotion, has been shown to improve trust, likeability and perceived intelligence in HRI [32]. The use of musical phrases to improve interaction in robotics still has many future avenues for research, in particular we believe group robotic environments, where multiple sources of speech may lead to higher cognitive loads and distraction, could be improved through musical phrases. The role of emotional contagion, where the emotion of a robot alters the emotion the a human collaborator has also not been addressed.

In this paper we explore how EMP can be used in HRI in a large group containing 10 robotic arms and one human collaborator (see Figure 11.1). We explore this area using an existing generator for emotional musical prosody [33] in combination with new custom gestures on a group of Ufactory Xarms. We aimed to explore the role of emotion contagion in a group of robotic arms performing a task, in this case moving an object between robotic arms that is passed from a human collaborator.

To develop these findings we conducted two studies, the first study was online through Amazon Mechanical Turk, and had 111 participants rating video footage of the robots interaction. We first analyzed whether EMP can lead to emotional contagion in human participants. We then considered the impact on HRI metrics for trust, competence, warmth and discomfort, as well as the relation between these metrics and the level of contagion. Our findings suggested that participants preferred robots using EMP, and emotional contagion could occur from robots to human participant. The second study was conducted with 16 participants in-person, primarily aiming to build on conclusions from the first study, while gathering extensive qualitative data to direct future research both using EMP, as well as music and HRI.

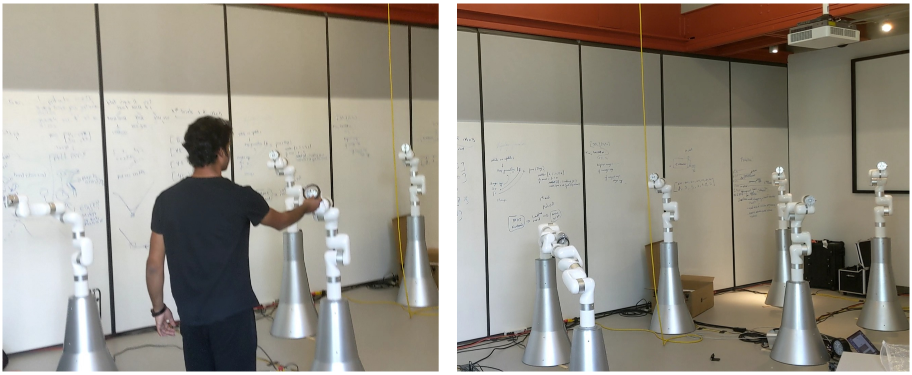

FIGURE 11.1
User passing ring to robot and robots passsing rings between each other.

11.2 Background

11.2.1 HRI and Emotion

Research in HRI often explores how different variables can alter human-based metrics. One of the most widely used survey in HRI is the Godspeed Questionnaire Series, which measures anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots [4,39]. Each metric in the Godspeed survey is measured with 4-5 bipolar sub-questions. Other surveys are often created for a specific metric with more extended questions such as ratings for self-efficacy [23] or willingness to interact [12]. It is also very common for psychology and social studies metrics to be adopted within the field of robotics, such as the mind attribution scale [16]. In HRI a reoccuring issue is the development of trust, as low levels of trust can lead to under-utilization in work and home environments [17]. Trust is generally categorized into either cognitive trust or affective trust [14]. Perceiving emotion is a crucial for the development of affective trust in human-to-human interaction [24], as it increases the willingness to collaborate and expand resources bought to the interactions [15].

Research focusing on the role of emotion in robotics has seen continual growth over the last thirty years, spanning many applications and platforms [31]. This research can primarily be divided into two main categories, emotion for improved performance (called “survivability”) and emotion for social interaction [2]. Survivability invokes the belief that emotion is central to animals’ ability to survive and navigate the world and can likewise be used in robots. This includes situations such as an internal emotion based on external danger to a robot [1]. The second category – social interaction – addresses anyway emotion is used to improve interaction, such as analyzing a humans emotion, or portraying emotion to improve agent likeability and believability [19].

11.2.2 Sound and HRI

There is only limited work in sound and HRI outside the use of speech systems, with research on the impact of sound relatively rare [42]. Studies have been conducted to analyze whether the use of a beep improves the perception of a robot with positive results, although more considered application of the range of possible audio sounds has not been conducted [11]. Consequential sounds are the sounds made by a robot in normal operation, such as motors and physical movement. The sound from motors has been used as a communication tool through modification of gesture [13], as well as used to improve localization [7]. Overall, consequential sounds have been analyzed for their impact on interaction with primarily negative results [18, 38]. Sonification of robotic movements has been examined, such as in relation to emotions for children with Autism Spectrum Disorder (ASD) [43], or for general movement of robots [5]. While there are multiple attempts to incorporate sound beyond spoken language into robotics, it is ultimately very limited in scope with broad potential for further research. There has not been the same sound tested on multiple platforms, or even the same platform in different interaction types, and each sound implementation is very rarely explored outside single one-off studies.

11.2.3 Emotional Musical Prosody

EMP was developed in previous work by the authors [29] to leverage the emotional power of musical improvisation, combined with the musical features of language, to create a new method of non-semantic communication called Emotional Musical Prosody (EMP). In linguistics, prosody refers to the parts of speech that are not the words, including the intonation, stress, and rhythm of the words. These features have parallels in music and the way a performer expresses music. EMP offers many advantages for HRI, by not using language, it can lower cognitive load while adding an emotion-driven personality, which has been repeatedly shown to improve collaboration metrics [33].

The first phase of developing EMP relied on gathering a new dataset from three vocalists, Mary Carter, Ella Meir and Aya Donahue. Emotions can be classified in countless ways, such as Ekman's six categories of anger, surprise, disgust, enjoyment, fear, and sadness or the cirumplex model, which places emotion on a two-dimensional circular space. We chose to use the Geneva Emotion Wheel (GEW), which uses 20 different emotion types, such as love and admiration. GEW allowed us to capture a wide range of emotions, within a range we felt would be possible for the performers. The final dataset contained 12 hours of audio, with around 450 phrases for each emotion. After collecting the dataset, the authors developed a new method using deep learning to generate audio phrases. This method focused on real-time generation, capable of rapidly responding to a human's emotion (more details on the underlying generation technology are available in [33]).

After the dataset and generation method were established, multiple studies aimed to understand how EMP could work in dyadic HRI (between one human and one robot) across multiple robotics platforms [10, 25–30, 33], as well as in Chapter 10 for portraying personality. EMP was compared between social robots, who are designed primarily for social interaction with humans, with a humanoid robot, which aims to replicate human behavior, and an industrial robotic arm [34]. EMP significantly improved trust, likeability and perceived intelligence for the industrial arm and social robot but not the humanoid robot. Emotional Contagion and EMP has not yet been explored in depth however, nor has large scale groups with interaction with human participants. Emotional contagion refers to the process of emotion transferring from one agent to another, commonly in the form of shared group laughter. In group human interactions, emotional contagion has been shown to improve cooperation and trust in team exercises [3, 21, 37]. Music contagion has been studied extensively, with a complex relationship developed between the emotion portrayed and the effect on music [8], nevertheless music has consistently lead to emotional contagion in many listeners [9].

11.3 Study 1: Fundamentals of Emotion Contagion, Music and Robotics

The primary goals of our first study were to engage a diverse audience, who would view groups of robots using EMP and establish some fundamental principles. Our goal was to understand participant perceptions of emotional contagion, rather than aiming to actually measure emotional contagion in the participants. We developed three research questions aiming to understand the role of emotional contagion, warmth, competence, discomfort and trust in task-basked group robotic activities. We analyzed the variation in robots when using EMP compared to performing a task alone, with tasks sometimes being performed successfully and sometimes failing.

- RQ1 Does EMP improve the ratings for warmth, competence, discomfort and trust compared to the task alone?
- RQ2 Does EMP lead to emotional contagion when compared to a task alone?
- RQ3 How do levels of emotional contagion compare to human ratings for warmth, competence, discomfort and trust?

Research Question 1 was designed to reconfirm that using EMP in a group of robots would improve the metrics of warmth, competence, discomfort and trust over the performance of a task alone. In previous work EMP has been measured in multiple use cases, but not in combination with a task with options for success or failure. Robots involved in specific tasks however is a much more real world scenario, than idle robots. It is possible that the addition of a task will override any emotional or other reaction that may have be drawn from prosody. We hypothesized however that ratings for warmth, competence and trust would all increase in the EMP version, with a lower rating for discomfort. We believed this would replicate previous studies reactions to EMP [34], despite the change of scenario and addition of success and failure in the task.

Research Question 2 explored whether the emotional reactions from robots would actually lead to participants believing a human user would show a different emotional response. We hypothesized that participants would be influenced by the emotional content of the audio, and change their results when compared to people who view the task alone, expanding on simpler previous findings about the role of interaction [28].

Research Question 3 aimed to identify how, if at all, the ratings for warmth, competence, discomfort and trust correlate with the likelihood a participant believing an emotional response would occur from the user. We conducted this question as an exploratory study, aiming to broadly see if any correlation existed. Our hypothesis was that emotional contagion would be linked to higher ratings of warmth, competence and trust, as participants would relate better to the robots when they identify with an emotional response.

11.3.1 Method

To answer these research questions we developed an online study, where participants would view video footage of the group of robotic arm. The stimuli for this experiment was a group of robots tasked with passing a ring from a human to a box. This task required interaction between the participant, robot, and group of robots. We used a ring passing task to ensure that the person's reactions to each emotion are a results of the emotion, and not from robots moving. The robots would fail to deliver the ring 50 % of the time. After the robot would succeed or fail in its task, the rest of the robotic group would react in response to the ring passers. If the task fails, the robots can react with either valance (happy or sad). If the task succeeds, the robots would have different arousal levels (happy or calm). This resulted in four possible conditions for EMP, either Task Failure Anger, Task Failure Joy, Task Success Calm, Task Success Joy. The robots emotional reactions were designed based on a set of emotional musical prosody phrases, previously validated [25] and described in section 11.2.3. The gestures were designed using the rule-based system described in Chapter 13.

Video samples can be seen at https://www.soundandrobotics.com/ch11

Participants first completed a consent form, followed by watching the video stimuli. The study was a within design, with all participants viewing every video. For each robot video participants rated the emotion a "human interacting with the robot will most likely feel". This was presented as a multiple choice question with the choices, "happy or excited", "angry or disgusted", "sad or depressed", and "relaxed or calm", with one option from each quadrant of the circumplex model. Participants also had an option to enter free text, or answer "None". Following the video we measured warmth, competence, and discomfort using The Robotic Social Attributes Scale (RoSAS) [6]. RoSAS is an 18 item scale that requires participants to identify how closely associated certain words are with the robot (such as "reliable" and "scary"). To measure trust we used the 14 point version of the Trust Perception Scale-HRI [36], which asks participants to rate the percentage of time a robot will act in a certain way, such as "dependable"; and "provide appropriate information". The questions for RoSAS and the Trust Perception Scale were randomly ordered for each participant. We also optionally asked for participant demographic information and allowed an open general text response to discuss the study, and for other comments on the robot.

The study was conducted online using Amazon Mechanical Turk (MTurk) to gather participants, with the survey hosted online using Qualtrics. We recruited 118 participants, of whom we used the responses of 111. Through the study there were a range of attention checks, including a video overdubbed with audio asking for a specific response, and a survey question requiring a specific response. We also tracked time spent on each question, and overall time on the survey, as well as participants IPs. In past Mturk studies we have found multiple participants working from the same IP, which is allowed on the platform, but prevents us from knowing if there was any collaboration. Overall, we recruited 118 participants, and we used the responses of 111 who all passed the attention checks. Participants were all based in the United States, with ages *(M = 44, STD = 10.4, max = 70, min = 20)* and 46 identify as female, 64 as male and one non-binary.

11.3.2 Results

11.3.2.1 RQ 1

For Research Question 1 we first calculated Cronbach's Alpha for the combined metrics in the trust survey. This gave a result of *0.91*, indicating high internal reliability for the questions. A pair-wise T-test showed did not find a significant result *(F=1.7, p =0.08)* after Holm–Bonferroni considering the four variables. The effect size, measured using Cohen's D was *0.32*.

For each question in RoSAS we first calculated Cronbach's Alpha, with the results: Competence *0.92,* Warmth *0.93*, and discomfort 0.81 indicating high internal reliability for each metric. A pair-wise indicated that competence was not significant, with an effect size of *0.06*. Warmth was significant *(F=5.5, p¡0.001* with an effect size of *0.94*. Discomfort was also significant, *(F3.6, p¡0.001* with an effect size of *0.61*. Figure 11.2 shows a box plot of the results.

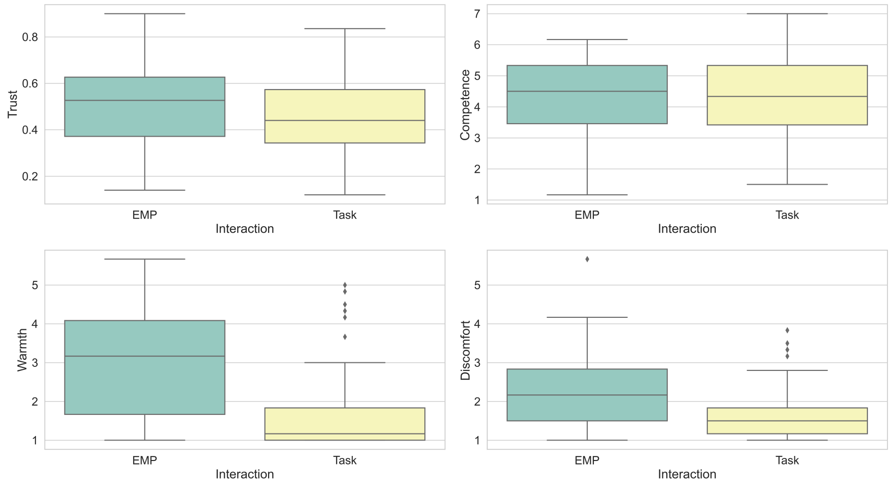

FIGURE 11.2
EMP compared to task for HRI metrics.

11.3.2.2 RQ 2

To analyze whether participants in an online study would recognize the potential for emotional contagion we compared responses between task alone and EMP conditions. Considering task failure, we found that with the addition of anger-tagged EMP participants, were much more likely to choose "sad or depressed" and likewise for task failure with Joy were much more likely to choose "happy or excited". For successful tasks we also found joy-tagged EMP increased the chance of choosing "happy or excited", however for success calm, participants were more likely to choose "sad or depressed". Figure 11.3 presents a bar plot of the results.

11.3.2.3 RQ 3

To answer Research Question 3 we conducted linear regression on each of the four metrics compared against the contagion rating. A result was as a positive contagion whenever the emotion matched the emotion of the robot. We found a significant result for Trust, Competence and Warmth, but not discomfort, indicating some correlation between each result and the level of contagion (see Table 11.1. Figure 11.4 displays a regression plot of the results.

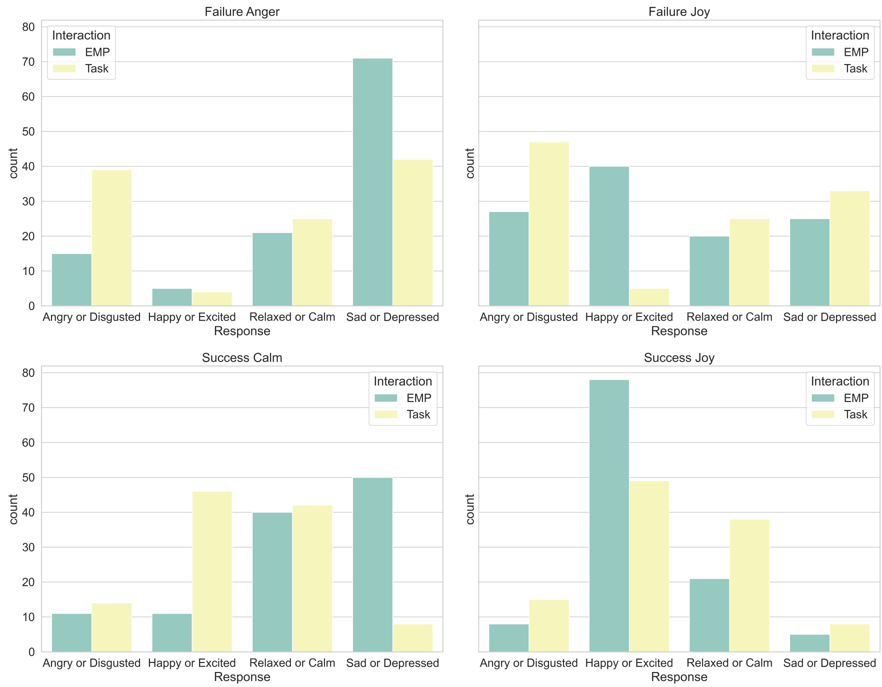

FIGURE 11.3
EMP compared to task for emotional response.

TABLE 11.1
Linear regression of interaction between emotion contagion and HRI metric.

	Slope	Intercept	r	p	Standard Error
Trust	0.065	0.446	0.339	0.011	0.024
Competence	0.351	3.945	0.265	0.048	0.174
Warmth	0.577	2.408	0.375	0.004	0.194
Discomfort	0.069	2.137	0.066	0.63	0.142

11.3.3 Discussion

Research Question 1 confirmed that results from past EMP studies were replicable and carried across to larger groups of robots in a different setting. We believe confirming these findings helps to strengthen past work in EMP, while justifying the continued exploration. Research Question 2 supported broad findings that emotion contagion could occur from robots human participants. We certainly make no claim that an online study in this manner can

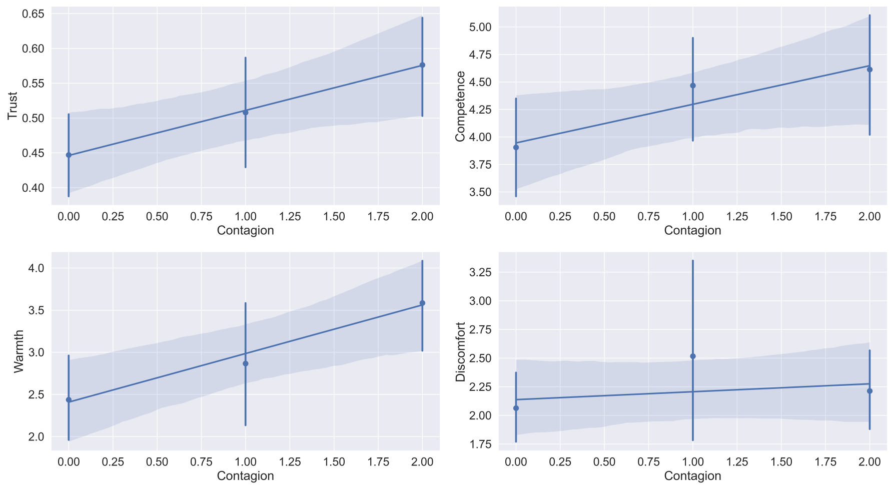

FIGURE 11.4
Correlation between emotion contagion and response.

accurately capture emotional contagion, however believe the participants believing emotional contagion would occur encourages future study in this direction. Research Question 3 further encouraged that the participants who believed there would be emotional contagion.

11.4 Study 2: Exploration Research into the Role of EMP in Groups

The results of our first study confirmed past studies and that EMP is able to change ratings for HRI Metrics. Additionally it suggested that emotional contagion could correlate with a robots emotional reaction. To continue exploring the role of EMP beyond past research we decided to conduct an in-person study with a participant interacting with many robots simultaneously. We were primarily interested in gathering a wide range of qualitative data that could help inform future understandings of EMP and sound in group robotics more broadly.

Our in-person study focused on an extended interaction with the robots that was video recorded, followed by a 20 minute semi-structured interview. Our stimuli includes three robots that people could interact with. Each robot had a unique color to help participants differentiate the three robots. In addition, participants had three sets of four rings (12 total). The sets of rings matched the colors on the robots. The colors on the rings signified a participant which robot to give the ring to. This ensures that each robot gets an even number of rings to interact with. After a robot receives a ring, it will pass it along to another robot, that places it in a box. Afterward, all the robots will react to the robots success/failure. The success and failure patterns were the same as the first study. We placed a camera in the group of robots that monitored the participants facial expressions and interactions.

As a participant first walked into the lab, they were asked to read and complete a consent form, that outlined the basic interaction with the robots and that they were being recorded on video. After signing the consent form, participants were given 12 rings and instructed to hand the top ring to the robot with a matching color. We instructed the users that the robots "Are tasked to pass the ring to a back robot, that must put the ring in a box". The participant continues to observe the robot pass the ring, place it into the box, and observe the robots reactions. While we did not initiate some information, we told participants that the robots were aware of task results if they asked. At the end of the 12 rings, participants were given 4 white rings that they were allowed to place on any robot. We documented the white ring placement. The task result and robot reactions are randomized for the white rings. The white rings served as base questions for the extended interview. Participants proceeded to complete a qualtrics survey about the robots and then spoke with a lab member for an extended interview.

Participants were undergraduate student volunteers at Georgia Institute of Technology. Each one was rewarded 20 dollars for participating. Of the 20 students who participated in the study, 16 results were used due to technical malfunctions in the environment.

After completing the study we conducted a constant comparison analysis on the results of the interviews to develop themes and ideas that emerged [20]. Constant comparison analysis begins by building categories and subcategories from the analyzed speech, then organizing codes and continual refinement of categorization. These themes split into three categories, firstly, perception of robot emotion, musical reactions and gesture, which focused on manners in which EMP shaped participant responses. The second category, robotic groups, featured responses specific to groups of robots. The third category HRI metrics, refers to comments from participants that relate to common HRI studies.

11.4.1 Perception of Robot Emotion, Musical Reactions and Gesture

11.4.1.1 Robot Emotion

During the study process we did not mention emotion at any stage. Nevertheless, the vast majority of participants interpreted the robots sound and movement as emotion-driven. This was not necessarily expected, as the embodiment of emotion into a robotic arm is rare [32] and no participants had interacted with arms previously. One participant described that the "Start was very unexpected" as they didn't expect an emotional response', while another participant echoed this general feeling, although also adding the "emotion was a bit scary at first, but then cool towards the end".

There was a general mixed perception about whether the gestures, the music, or a combination of both was driving the perception of emotion. Some descriptions of how the robots conveyed emotion focused entirely on the gesture, such as, "looking down and not making eye contact is kind of considered to be sad and just looking up and jumping around is happier", "I thought it was sad is because it was shrinking into a corner and turning away from me and "when it was like more like pessimistic they would like bow their heads more and like it was like kind of it felt kind of dispirited". However other participants believed the gestures alone did not convey emotion, "but the audio made it an emotion, otherwise it will just be like just arms moving up and down". Descriptions of the emotional content of music included "it was music designed for robots to specific emotions" and that the music "showed characteristic things humans associate with emotions".

Participants generally interpreted the emotions as binary, primarily as happy or sad but also used terms such as "more optimistic or more pessimistic" or "positive" and "negative". While the robots were designed to show four emotions, no participant consciously noted a difference between the level of emotion displayed by the robots. One participant did disagree with the notion of including emotion in a robotic system, stating "they don't if they should be happy, so couldn't draw emotion from anything". No other participants however questioned the reason for including emotional responses.

11.4.1.2 Musical Reaction

There was fairly limited reaction to the music as a separate feature of the robots, outside of the emotional role. Participants agreed that the music showed emotions, but only one participant commented on the aesthetic decisions of the music. One participant notes that it was interesting that the robots were using a human-like voice, but they believed it fit the robot. Ultimately all participants viewed the music as an embedded part of the robotic system, and generally not a separate process.

11.4.1.3 Gesture

Many participants commented on the auxiliary gestures of the robots, and it was common for the participants to be intrigued by the relations of the music to gesture. "I felt like that was some kind of like abstract dance, but it was also cool to see them like move around in different directions". No participants questioned the utility of the additional gestures, as we had expected may occur, with participants instead focusing on understanding what the gestures where trying to convey, as described in Section 11.4.1.1.

11.4.1.4 Language Choices

Participants often questioned their own choice of phrases and language in describing the robot. We did not give the robot any name, or imply any gender during the studies, allowing the participants to develop their own understanding of the platform. Multiple participants paused when assigning a gender to the robot, including changing direction mid-sentence, such as: "It was dropped by him – I don't know why I'm calling in to him, but it dropped it and they were happy anyway". Throughout the interviews other participants would use phrases such as "the robot tried" and "the robot wanted" before then self-correcting to remove the agency from the robot.

While many participants did not their use of language in describing robots, others comfortably personified the robots. Statements included language such as, "It is sometimes like one started partying and the other was like joining in the party and the third is like, oh, I see, uh party, let me join in and then finally it was just the last one partying" or "Sometimes it was just funny that the board dropped it and it just was like, yeah, congratulations or something like that". Many participants also independently described the robots as animals, such as an "octopus" or "spider". Other participants note that they seemed like "a weird animal", or "a group of a foreign species".

11.4.2 Robotic Groups

11.4.2.1 Group Appearance and Interaction

It was extremely common for participants to be taken aback by the group size and number of robots in the room. None of our participants had previously interacted or seen a single robot arm in person, possibly adding to the sometimes jarring experience of encountering a group of robots. Participants noted that they were "initially confused by why there were so many" and "very surprised by look of so many robots". Nevertheless, by the end of the experiment, and after interacting with the robots, all participants noted they felt comfortable with the group of robotic arms.

They were multiple perspectives on the concept of entitativity arising from the interviews, with participants generally divided about whether the arms were a single entity or a group of individuals. One participant described the arms as a robot "passing between itself", while others saw the robots as "10 robots all doing their own thing". Another participant noted "So like they yeah they function together apart like rather than having their like individual responses". A different participant noted that there were "three leader robots". Ultimately the perception of the group was extremely varied, ranging across many ideas without any clear conclusions between all participants.

11.4.2.2 Group Interaction

The form of interaction amongst the group was also widely discussed by participants. There were many references to group coordination and synchronization such as "I do think the group coordinated, but like they each had their own little changes", "a lot of them were synchronized as well" and "the robots did coordinate as a group, which was like memorable like they would do the same dance movements and they would have like their tunes would harmonize with each other, which was cool".

11.4.3 HRI Metrics

11.4.3.1 Self-Efficacy

An important component of the interviews was developing an understanding of participants self-efficacy with the group of robots. Self-efficacy is an individuals confidence that they will be able to successfully interact with a robot, and has implications for how often people want to interact with a robot and the sort of tasks they are comfortable completing [23]. From our interviews all participants felt "very confident in my own tasks", with no participants mentioning any concern over their ability to interact with the robots or to recognize how they could interact.

11.4.3.2 Trust and Confidence

In comparison to self-efficacy, trust and confidence refers to the users perception that the robot will behave reasonably, while self-efficacy refers to the humans' confidence in their own role. We found multiple lines of thoughts amongst the participants. A small minority stated their confidence and trust as a binary, either believing the system is trustworthy or not. The majority of participants stated they trust the robot to some degree but would be uncomfortable with tasks, with statements such as, "if it was like a hot beverages I wouldn't use it. I would be like scared".

Additionally most participants asked for extra information before they would be confident and trust the robots in a wider range of circumstances. This included requests for more details on how exactly the robots worked, to help improve understandings of when the robots were performing as expected. Statements include "I'd need to understand how they work better to use them more, but comfortable overall and think I would be happy to use". Multiple participants also noted that would be confident if they robots underwent some further troubleshooting.

Lastly, it is worth mentioning that one participant consciously noted that they felt "safe as I believed it was a controlled experiment". As for any study, the effect of environment should be considered before assuming real-world implications.

11.4.3.3 Intimidation

In our interview process, no participants described any level of intimidation toward the arms. One participant noted, "robotic arms are not very scary as such because they can't move other than the rotation, so I wasn't very intimidated, intimidated about". We believe this may however be a direct result of our participant pool, who were all undergraduate students at ANON. While no participants had experience with robotic arms before, considering the general university environment they are much less likely to be intimidated by the robots than a general population.

11.4.3.4 Reasons for Failure

When asked why the robots were failing the vast majority of participants assigned the error to human programmers and not the robot itself. Participants went as far as to suggest the human programmer "miscalculated the angles" or that the task was "incorrectly designed for the robot". Many participants also noted they were very surprised when the robot first dropped a ring, and then assumed there was a bug that led to the errors, instead of independent robot errors.

11.5 Discussion

Throughout both studies we found multiple areas arise that are worthy of further study and description. We contend that our results in previous sections imply a vast range of potential interactions between musical study and HRI.

11.5.1 Memorable Moments around Musical Interaction

A key idea that occurred through the interview process was that the most memorable and interesting moments were based on musical interactions. From the participants point of view these interactions were not always about the music, but could be what the music implied through the robot platform. For example, a participant noted their favorite moment was "When they [the robots] laughed at one of the robots failing". Even in a such a unique setting where a participant was seeing a group of many robots for the first time, music has the potential to enhance and change the interaction.

11.5.2 Music Adds Emotion

We found that music was clearly able to add a perceived emotion to the robot systems. While in some ways this is an expected result, the importance of almost unanimous description of robotic emotion from the participants should not be overlooked. Robotic arms themselves are inherently non-emotional, and in robotics literature are very rarely used to display emotion. In this way, music can add an entire new range of approaches to interacting with this technology.

11.5.3 Embodied Music

In our setup, music was coming directly from underneath each robot, with each robot having it's own speaker. From all feedback it was clear that participants heard the music as from each robot itself, and not as a generalised sound. In early prototypes we had considered having four speakers or other arrangements, but chose to use a speaker in each arm, despite the extra system complexity. Before the study there was the potential for the musical phrases to not be directly associated with the robot and instead as background music. While we did not test this directly, we believe speaker placement, where it was embodied within the robot, was important to the perception of the voices as coming directly from each robot and drastically alters the interaction.

11.5.4 Music in Groups

We found very mixed perspectives on the role of music in groups and how it altered the perception of the group dynamics. There is a range of future work that could be done in this area. One potential direction is more variability in the timbre of the sound, or trying different melodic approaches for each robot. For this initial research we consciously choose to reduce the variables and use the same voice for each robot.

11.5.5 Music and HRI

The intersection of musical phrases and HRI is very much an under researched area, with only minimal work addressing the areas. With this in mind, it is worth continuing to describe the important role that music could have. We confirmed in our first study that the addition of music can increase trust and likeability. Music however could foreseeably have many other roles in robotics, such as extra dissonance reducing trust when a system should be avoided. This work is ultimately only one of many potential musical approaches to HRI.

11.5.6 Limitations

Both the studies presented in this paper were carefully designed to avoid limitations when possible. However, as in any study there limitations within each study. The online studied used pre-recorded videos instead of live interaction. We believe that for this was an acceptable experimental design as ultimately and allowed us to initially gather a wide range of data from a very diverse participant pool. Multiple past papers have shown no significant variation in results when a participant is watching a robot on video compared to in person [40,41]. By combining an online and in-person study we aimed to collect a very diverse range of opinions on robotic systems and develop a foundation for future research.

A further limitation was the length of each study, which was conducted over a single session. One participant even noted that "I mean, if you're doing that the whole day everyday, like I mean that's gonna get boring". In the future we aim to study longer interactions over multiple sessions to identify the variation that occurs during repeat encounters.

11.6 Conclusion

This research reiterates the important role music can have in communication and HRI. We were able to demonstrate that EMP is capable of changing HRI metrics in an online study, confirming past results, while suggesting the possibility of emotional contagion. Our second study focused on broad qualitative results, aiming to emphasize the perception of music in robotic systems and develop further avenues for research. We intentionally focused on robotic arms as this is both a world-wide growth area but also a platform well suited to gain from the addition of music. To robotic arms, music can add emotion and an entire range of communication options. We believe this form of interaction can enhance collaboration across many settings, ranging from robotic arm interface testing, to large scale factory use. Overall, we hope this research helps expand broader ideas about the possibilities of sound and robotics

Bibliography

[1] Arkin, R. C., Fujita, M., Takagi, T., and Hasegawa, R. An ethological and emotional basis for human–robot interaction. *Robotics and Autonomous Systems 42*, 3-4 (2003), 191–201.

[2] Arkin, R. C., and Ulam, P. An ethical adaptor: Behavioral modification derived from moral emotions. In *2009 IEEE International Symposium on Computational Intelligence in Robotics and Automation-(CIRA)* (2009), IEEE, pp. 381–387.

[3] Barsade, S. G. The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly 47*, 4 (2002), 644–675.

[4] Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics 1*, 1 (2009), 71–81.

[5] Bellona, J., Bai, L., Dahl, L., and LaViers, A. Empirically informed sound synthesis application for enhancing the perception of expressive robotic movement. Georgia Institute of Technology.

[6] Carpinella, C. M., Wyman, A. B., Perez, M. A., and Stroessner, S. J. 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.

[7] Cha, E., Fitter, N. T., Kim, Y., Fong, T., and Matarić, M. J. Effects of robot sound on auditory localization in human-robot collaboration. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (2018), pp. 434–442.

[8] Davies, S. Infectious music: Music-listener emotional contagion.

[9] Davies, S., Cochrane, T., Scherer, K., and Fantini, B. Music-to-listener emotional contagion. *The Emotional Power of Music: Multidisciplinary Perspectives on Musical Arousal, Expression, and Social Control* (2013), 169–176.

[10] Farris, N., Model, B., Savery, R., and Weinberg, G. Musical prosody-driven emotion classification: Interpreting vocalists portrayal of emotions through machine learning. In *18th Sound and Music Computing Conference* (2021).

[11] Fischer, K., Lohan, K., Saunders, J., Nehaniv, C., Wrede, B., and Rohlfing, K. The impact of the contingency of robot feedback on hri. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on* (2013), IEEE, pp. 210–217.

[12] Fraune, M. R., Nishiwaki, Y., Sabanović, S., Smith, E. R., and Okada, M. Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (2017), pp. 205–213.

[13] Frederiksen, M. R., and Stoey, K. Augmenting the audio-based expression modality of a non-affective robot. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)* (2019), IEEE, pp. 144–149.

[14] Freedy, A., DeVisser, E., Weltman, G., and Coeyman, N. Measurement of trust in human-robot collaboration. In *Collaborative Technologies and Systems, 2007. CTS 2007. International Symposium on* (2007), IEEE, pp. 106–114.

[15] Gompei, T., and Umemuro, H. Factors and development of cognitive and affective trust on social robots. In *International Conference on Social Robotics* (2018), Springer, pp. 45–54.

[16] Kozak, M. N., Marsh, A. A., and Wegner, D. M. What do i think you're doing? action identification and mind attribution. *Journal of Personality and Social Psychology 90*, 4 (2006), 543.

[17] Lee, J. D., and See, K. A. Trust in automation: Designing for appropriate reliance. *Human Factors 46*, 1 (2004), 50–80.

[18] Moore, D., Martelaro, N., Ju, W., and Tennent, H. Making noise intentional: A study of servo sound perception. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI* (2017), IEEE, pp. 12–21.

[19] Ogata, T., and Sugano, S. Emotional communication robot: Wamoeba-2r emotion model and evaluation experiments. In *Proceedings of the International Conference on Humanoid Robots* (2000).

[20] Onwuegbuzie, A. J., and Frels, R. *Seven Steps to a Comprehensive Literature Review: A Multimodal and Cultural Approach*. Sage, 2016.

[21] Ososky, S., Schuster, D., Phillips, E., and Jentsch, F. G. Building appropriate trust in human-robot teams. In *2013 AAAI Spring Symposium Series* (2013).

[22] Pelikan, H., Robinson, F. A., Keevallik, L., Velonaki, M., Broth, M., and Bown, O. Sound in human-robot interaction. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA, 2021), HRI '21 Companion, Association for Computing Machinery, pp. 706–708.

[23] Robinson, N. L., Hicks, T.-N., Suddrey, G., and Kavanagh, D. J. The robot self-efficacy scale: Robot self-efficacy, likability and willingness to interact increases after a robot-delivered tutorial. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (2020), IEEE, pp. 272–277.

[24] Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. Not so different after all: A cross-discipline view of trust. *Academy of Management Review 23*, 3 (1998), 393–404.

[25] Savery, R. Emotional musical prosody: Validated vocal dataset for human robot interaction. In *2020 Joint Conference on AI Music Creativity,* (2020).

[26] Savery, R. *Machine Learning Driven Emotional Musical Prosody for Human-Robot Interaction.* PhD thesis, Georgia Institute of Technology, 2021.

[27] Savery, R. Machine learning driven musical improvisation for mechanomorphic human-robot interaction. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (2021), pp. 559–561.

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

[29] Savery, R., Rose, R., and Weinberg, G. 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)* (2019), IEEE, pp. 1–7.

[30] Savery, R., Rose, R., and Weinberg, G. Finding shimi's voice: fostering human-robot communication with music and a nvidia jetson tx2. In *Proceedings of the 17th Linux Audio Conference* (2019), p. 5.

[31] Savery, R., and Weinberg, G. 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).

[32] Savery, R., and Weinberg, G. Robots and emotion: a survey of trends, classifications, and forms of interaction. *Advanced Robotics 35*, 17 (2021), 1030–1042.

[33] Savery, R., Zahray, L., and Weinberg, G. Before, between, and after: Enriching robot communication surrounding collaborative creative activities. *Frontiers in Robotics and AI 8* (2021), 116.

[34] Savery, R., Zahray, L., and Weinberg, G. Emotional musical prosody for the enhancement of trust: Audio design for robotic arm communication. *Paladyn, Journal of Behavioral Robotics 12*, 1 (2021), 454–467.

[35] Savery, R., Zahray, L., and Weinberg, G. Shimon sings-robotic musicianship finds its voice. In *Handbook of Artificial Intelligence for Music*. Springer, Cham, 2021, pp. 823–847.

[36] Schaefer, K. E. 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.

[37] Stock, R. M. Emotion transfer from frontline social robots to human customers during service encounters: Testing an artificial emotional contagion modell.

[38] Tennent, H., Moore, D., Jung, M., and Ju, W. Good vibrations: How consequential sounds affect perception of robotic arms. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2017), IEEE, pp. 928–935.

[39] Weiss, A., and Bartneck, C. Meta analysis of the usage of the godspeed questionnaire series. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2015), IEEE, pp. 381–388.

[40] Woods, S., Walters, M., Kheng Lee Koay, and Dautenhahn, K. Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In *9th IEEE International Workshop on Advanced Motion Control, 2006.* (2006), pp. 750–755.

[41] Woods, S., Walters, M., Koay, K. L., and Dautenhahn, K. Comparing human robot interaction scenarios using live and video based methods: towards a novel methodological approach. In *9th IEEE International Workshop on Advanced Motion Control, 2006.* (2006), IEEE, pp. 750–755.

[42] Yilmazyildiz, S., Read, R., Belpeame, T., and Verhelst, W. Review of semantic-free utterances in social human–robot interaction. *International Journal of Human-Computer Interaction 32*, 1 (2016), 63–85.

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