Emotion Musical Prosody for Robotic Groups and Entitativity

Richard Savery, Amit Rogel and Gil Weinberg[1]

Abstract— **Research in human-robot interaction has focused on the relationship between a single robot and a single human participant. Only limited research has addressed the contrasting dynamic when humans interact with a group of robots. This dynamic adds additional human-robot interaction considerations, such as the level of entitativity, which is the identification of a group as a single entity as opposed to a collection of individuals. This paper proposes that emotional music prosody can play a key role in improving the interaction between humans and groups of robots by modifying the level of entitativity. Musical prosody refers to the use of pitch, rhythm and timbre features derived from language, but used without semantic meaning.**

We conducted a between-group experiment, presenting to subjects a group of industrial robotic arms performing a task either without sound, with the same emotional musical prosody voice for each robot, or with contrasting voices for different robots. We were able to show with significant results that the use of musical prosody improved likeability and trust over soundless gestures for groups of robots. We also demonstrate that, through subtle variations, prosody is able to alter the level of entitativity perceived by external observers. Finally, our results indicate a complex relationship between entitativity and common HRI metrics with higher levels of entitativity leading to improved performance, contradicting past literature.

I. Introduction

The vast majority of research in Human-Robot Interaction (HRI), focuses on interactions between a single human and single robot participant. Studies that have been conducted on group interaction show differences in the perception of a robot in a group, compared to individually. These include willingness to interact as well as levels of fear [1]. These issues are often exaggerated for non-anthropomorphic robots in groups, with results indicating that such robots are more threatening and less likely to encourage human engagement [2].

A key issue with groups of robots is the amount of entitativity perceived by human collaborators. Entitativity refers to the level in which a group is seen as a single entity, such as multiple arms being viewed as a single robot, compared to individual agents. Understanding entitativity in human interaction is considered crucial for developing fundamental understandings of human group dynamics [3]. It has been demonstrated that the perception of higher levels of entitativity will create a negative image of the group with less chance of external interaction [4]. For robots, entitativity has only recently entered consideration, with some findings linking higher entitativity to a reduced perception of friendliness

Fig. 1. uFactory's xArm: a 7 degree of freedom industrial arm dropping rings in a box

and comfort while increasing ratings for "unnervingness" and "creepiness" [5].

In this work, we show that emotional musical prosody can serve multiple purposes in group robot interaction. Emotional musical prosody involves leveraging the non-linguistic features of speech, such as pitch, rhythm and timbre as a form of communication. We contend that emotional musical prosody is effective because the portrayal of an emotional agent is crucial for collaboration and creating a believable computer agent [6]. Emotional prosody can also help bypass the uncanny valley - where a robot becomes unappealing due to attempting to appear human-like [7] - by communicating in a mechanomorphic manner rather than trying to replicate human speech. Musical prosody has been shown to be effective in single robot interaction over gesture and text-to-speech, by increasing the trust for social robots [8] and likeability for industrial arms [9].

We believe these benefits can be extended from individual robots to groups of robots, improving key metrics for industrial arms. We also contend that as emotional musical prosody can be easily modified with timbre shifts, it can support reducing entitativity. Such changes to the sound of the prosody can imply variation between robots in a group setting, allowing an easy format to reduce entitativity.

We conducted a between-groups experiment, comparing three industrial arms performing a collaborative task with a human participant. Participants were shown either the arms without prosody, each arm performing with the same prosody, or the arms performing with variations of the same prosody. We found significant improvements for trust

*This material is based upon work supported by the National Science Foundation under Grant No. 1925178

[1]Georgia Tech Center for Music Technology, Atlanta, GA, USA, rsavery3@gatech.edu

and likeability for the prosody robots, with no variation for participants willingness to interact, confidence with the system or perceived intelligence. Arms performance with different versions of prosody had the lowest rating for entitativity, while using the same prosody achieved higher ratings for entitativity. We also examined the relationship between entitativity ratings and each HRI metric such as trust and likeability, and found that higher levels of entitativity lead to increased ratings across all metrics, contradicting past findings [2].

II. Background

Robotic pose, gestures, and facial expressions have been used to communicate mood [10], [11], emotional state[12], and intent [13]. Ende et al. studied the recognition of typical human gestures performed by robots. Communicative gestures such as "stop" and "move over here" were studied and compared between an industrial arm and a humanoid robot. Humans found it more difficult to recognize the nonverbal cues of industrialized arms over humanoid structures [13]. Researchers have also used gestures to improve likeability [14], perception of animacy [15], and trust between humans and robots [16].

While gestural methods of communication have been successful at reducing stress and improving perceptions of robots, they cannot support subtle and non distracting communications [17]. Using gestures is the most common nonverbal communication method for humans [18], however gesture use in industrial arms can overlap with their primary functions and safety tasks [19]. In addition, these modes of interaction have been difficult to upscale to larger groups of robots, partly because as robots increase in quantity, humans become more intimidated and have a negative response towards the robots [20].

One approach to provide non distracting communication for robots is using sound [14], [21]. More specifically, it has been shown that non vocal prosody [22] can help communicate important background information to humans. In our work, we developed a novel music-based approach for robotic prosody [8], [9] which has improved the perception of animacy, trust, and likeability and changed the overall ratings for key robotic tasks [23], [24]. As industrial arms are integrated more into society, the quantity of arms in groups will increase [25], introducing a novel research domain of robotic entitativity.

Entitativity measures how a group is perceived as a coherent unit rather than separate individuals [26]. In psychology, it is used as an important measurement for group dynamics and effectiveness. Castano theorizes that four main factors impact a group entitativity: common fate, similarity, salience, and boundedness [27]. In human groups, people can relate more to high entitativity groups than low entitativity groups [28], [4]. Hamilton suggests that outsiders are more likely to engage in integrative processing of groups with high entitativity [29]. Increased entitativity will also increase the perceived unification of the group. In human groups, high entitativity requires increased coordination and focus on unification to accomplish a task [29].

While human entitativity has been widely researched, there have been limited studies on the perception of entitativity in robotic groups. Fraune found that increasing the quantity of robots would create more negative emotions towards the robots. A higher quantity would increase anxiety and fear levels of humans [30], [2]. Abrams showed synchronicity in robot movements can vary entitativity and appear scary to an observer. However, robots that appear unique would leave a warmer impression, and increase the desire to work with humans [20]. Saunderson found that a large amount of robots in groups can negatively impact a human's impression and trust [10]. We believe that the work described in this paper on integrating music driven emotional prosody into robotic groups can address and mitigate this negative effect of robotic entitativity.

III. Method

We investigate three research questions to study the intersection of robots in groups, entitativity and musical prosody:

- RQ 1 Can emotional musical prosody improve Likeability, Perceived Intelligence, Trust, Confidence and Willingness to Interact, for a group of robots?
- RQ 2 Can variations in emotional musical prosody lower the level of entitativity for a group of robots?
- RQ 3 How does the level of entitativity correlate with Likeability, Perceived Intelligence, Trust, Confidence and Willingness to interact?

Research question 1 focuses on understanding the relationship between common HRI metrics and groups of robots. For this question, we are only interested in comparing the same prosodic voice for each robot against gestures, with the goal of replicating improvements shown in past studies with individual robots. The metrics were chosen due to past use in both group studies [2], [1] and studies with individual robots and prosody [8], [9]. Our hypothesis is that each metric will be improved by prosody with a significant result, replicating the results that have occurred for individual robots.

Research question 2 aims to compare the level of entitativity between three groups, one with gestures alone, one with a single voice and one with variations on prosody. Our hypothesis is that the single voice and gesture will perform similarly, while the multiple voices will achieve a lower level of entitativity, implying the appearance of multiple agents in the group.

Research question 3 is an exploratory question, designed to identify the relationship between entitativity and each metrics. We believe that higher levels of entitativity will correlate with reduced metrics as supported by research in human psychology and past research in HRI.

A. Measures

For each metric we used either an established measure or a combination of existing measures. To measure likeability and perceived intelligence we used a subset of the widely used Godspeed survey [31]. Participants were asked to rate

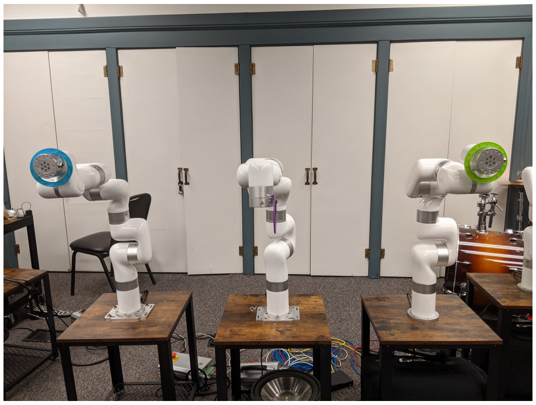

Fig. 2. Three xArms used for the stimuli. Each xArm was tasked to transport a ring to a box behind them

their impression of likeability and perceived intelligence for five questions on a scale of 1-5. We measured willingness to interact and confidence to interact each with three questions on a Likert scale, combined from past surveys [32], [30], [2]. To measure trust we used Schaefer's 14-point scale with participants rating each question from 0-100% to give a total trust percentage. To the common survey answers we added a "Not Applicable" option, as suggested by Chita-Tegmark et al. [33] to allow participants to avoid responding to aspects of trust they feel are not applicable to the industrial arms. We collected participant's age, identified gender and country of origin.

There is no standard accepted measure of entitativity, with HRI studies commonly combining multiple metrics from social studies, psychology and other HRI papers [34], [30]. Common questions range from defining entitativity for the participants and then asking directly for a rating [35], to attempts to combine other metrics such as friendliness, creepiness, comfort and unnerving into a rating [5]. We chose to measure entitativity using the survey proposed and validated by Blanchard et al. [3] which was shown to be effective for online and in person analysis. This measure consisted of three questions on a 7-point Likert scale.

B. Stimuli

We used a two minute video as our stimuli, overdubbed with different audio for each group. The video showed three robotic arms (shown in Figure 2) interacting with a human user. The human user placed a ring on each arm, which the robot then placed in a box behind itself. Each robot used the same gestures to place the rings in the box. The robot used in the study was an xArm, a 7 degree of freedom industrial arm made by uFactory.

We created three versions of the video with different audio, starting with a gesture only version which did not have any added audio. To add audio for other two videos we used prosody phrases from an existing dataset [36]. From the dataset we chose the emotions tagged as admiration, contentment, and compassion, each low arousal high valence emotions. We chose these emotions as we believed they best matched the interaction environment in the study, and using low valence emotions, such as sadness or anger would not be appropriate for this particular interaction.

The second version of the video used a matching voice (referred henceforth as single voice) for each robot. For each interaction the single voice used a different prosodic phrase, but had matching timbre, essentially sounding like the same voice singing a different phrase each time. For the third version of the video we used three different versions of the voice from the dataset. We also added variations to each voice through pitch shifting, a formant filter and modulation. This had the effect of sounding like three different voices, one for each robot. All three versions maintained the room sound and sounds of the robots movements. All stimuli can be viewed online. [1]

C. Participants

We recruited 60 participants on Prolific and 108 participants on Amazon Mechanical Turk (MTurk) to complete the study. Each participant was paid $2.00. We selected only MTurk Masters to participate and had no restrictions on prolific. We used multiple attention checks to verify each participant, and disqualified any data that failed any check. Our first attention check consisted of a spoken phrase at the end of the video requesting participants to type a random word on the next screen. We also had a question in the trust survey requiring participants to choose 10%. In addition to direct questions, we tracked the time spent on each question and the video, with any participant who did not watch the entire video removed. Finally, we removed two participants who completed the survey a second time, we assume after realizing they missed the audio from the attention check and restarting. From Prolific 6 participants failed an attention check, while 9 on MTurk failed an attention check, leaving us with a total of 153 participants.

In total we had 49 participants in the gesture only group, 48 in single prosody and 56 in the multiple audio. Participants place of origin was spread across 22 countries, with the majority from United States of America (n=71), India (n=22), Poland (n=14), Portugal (n=11), Mexico (n=8) with the remaining countries each have 5 or less participants. We found no significant variation in responses from each country, with the countries with less than 5 each fitting within the range of majority of responses. We had 62 participants identify as female and 90 as male, also with no significant variation between groups. The mean age of participants was 37 with a standard deviation of 12 and ranging from 18 to 75.

D. Protocol

The survey was conducted online using Qualtrics. Participants first completed a consent form and entered their MTurk or Prolific ID to indicate consent. They were then given instructions to watch the stimuli video with headphones connected. Participants were randomly assigned to one of the

[1] www.richardsavery.com/prosodyentatitivitystudy

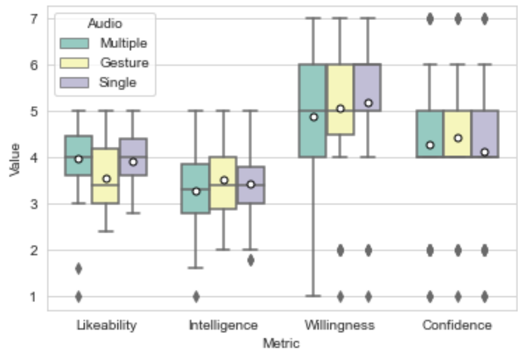

Fig. 3. Box Plot of Likeability, Intelligence, Willingness and Confidence

three groups of the study. Following the video participants first entered the text for the attention check and then completed the previously described measures. The measures were randomly ordered for each participant, with the sub-questions (such as each component of the trust survey) also randomly ordered for each participants. After completing each measure participants entered their demographic details and had a open text field with a prompt asking for any feedback on the robot system or experiment in general.

IV. Results

A. RQ 1: HRI Metrics

1) Likeability and Perceived Intelligence: The Cronbach's Alpha results for Likeability and Perceived Intelligence were 0.869 and 0.866 respectively, indicating high internal reliability for both measures. Perceived Intelligence had the results for single voice (mean = 3.324, std = 0.716, effect size = 0.050), multiple voices (mean = 3.271, std = 0.880, effect size = 0.230) and the gestures alone (mean = 3.527, std = 0.764, effect size = 0.240), with effect size calculated using Cohen's D. We ran a one-way ANOVA with the result $p > 0.05$, indicating the result was not significant. Perceived intelligence did not have a significant different between groups with each category having similar means and standard deviations, which did not support our hypothesis.

Likeability had the results for single voice (mean = 3.931, std = 0.600, effect size = 0.246), multiple voices (mean = 3.975, std = 0.800, effect size = 0.285) and the gestures alone (mean = 3.553, std = 0.736, effect size = 0.545), with effect size calculated using Cohen's D. We ran a one-way ANOVA with the result p = 0.007, indicating the result was significant. Likeability was improved significantly for both versions of prosody over the gestures alone, supporting our hypothesis. Figure 3 shows a box plot of the results for likeability and perceived intelligence. Perceived intelligence did not have a significant different between groups with each category having similar means and standard deviations, which did not support our hypothesis.

2) Confidence and Willingness: Confidence had the results for single voice (mean = 4.142, std = 1.607, effect size = 0.128), multiple voices (mean = 4.285, std = 1.637, effect size = 0.003) and the gestures alone (mean = 4.42, std = 1.363, effect size = 0.151), with effect size calculated using Cohen's D. We ran a one-way ANOVA with the result $p > 0.05$, indicating the result was not significant. Willingness had the results for single voice (mean = 5.183, std = 1.409, effect size = 0.158), multiple voices (mean = 4.875, std = 1.663, effect size = 0.150) and the gestures alone (mean = 5.064, std = 1.699, effect size = 0.026), with effect size calculated using Cohen's D. We ran a one-way ANOVA with the result $p > 0.05$, indicating the result was not significant. Neither confidence or willingness showed a significant result, indicating that prosody did not improve either of these metrics. Figure 3 shows a box plot of these results.

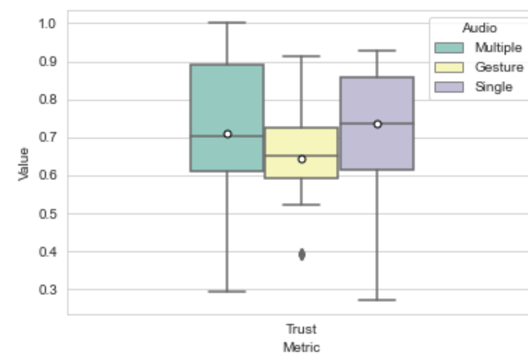

Fig. 4. Box Plot of Trust Ratings

3) Trust: To analyze our trust results we first calculated Cronbach's alpha which gave the result of 0.859, indicating high internal reliability. For single voice the results were (mean = 0.734, std = 0.146, effect size = 0.376), multiple voices (mean = 0.710, std = 0.166, effect size = 0.125) and the gestures alone (mean = 0.642, std = 0.710, effect size = 0.592), with effect size calculated using Cohen's D. We ran a one-way ANOVA with the result p = 0.009, indicating the result was significant. This supported our hypothesis that prosody would increase trust over gesture. Figure 4 shows the results as a box plot.

B. RQ 2: Entitativity and Prosody

For the three entitativity questions we first calculated Cronbach's Alpha, which gave a result of 0.88, indicating high internal reliability across the questions. For gestures alone the results were (mean = 4.241, std = 1.699), the single voice (mean = 4.490, std = 1.667) and multiple robots (mean = 3.601, std = 1.706). A one-way ANOVA gave a p-value of 0.022 indicating the results was significant. Additionally the multiple voices had an effect size calculated with Cohen's D of 0.45, indicating a medium effect size. This supported our hypothesis that subtle variations in voice would increase the entitativity of the group. Figure 5 shows a box plot of the results.

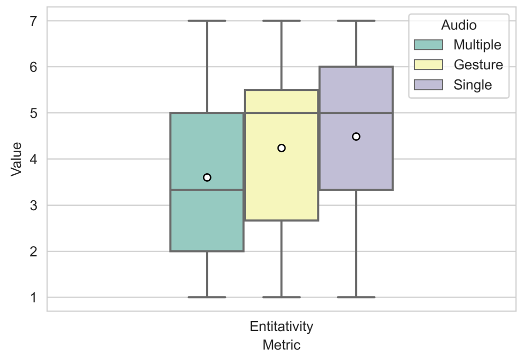

Fig. 5. Box Plot of Entitativity Ratings

C. RQ 3: Entitativity and HRI Metrics

For research question 3 we fit a linear regression model for each metric with entitativity. Table I shows the slope, intercept, r, p and error for each metric. Each metric tested had a positive slope, with higher levels of entitativity correlating with higher ratings. This did not support our hypothesis, as we had expected the opposite to occur across every metric.

	Slope	Intercept	r	p	Error
Willingness	0.294	3.64	0.286	p <.001	0.081
Intelligence	0.138	2.79	0.29	p <.001	0.037
Trust	0.025	0.567	0.31	p <.001	0.006
Likeability	0.075	3.313	0.162	0.046	0.037
Confidence	0.308	2.888	0.327	p <.001	0.073

TABLE I
LINEAR REGRESSION STATISTICS

V. DISCUSSION

A. RQ 1: HRI Metrics

We found that embedding prosodic sound to accompany co-bot arms gestures improved human's trust and likeability for these robots with significant results. Since in previous work, prosody improved trust and likeability in individual robots, it was expected that the improvements would carry across to groups. These metrics supports one of the core principles behind the use of emotional prosody in robots, namely that by increasing a robot's presence as an engaging emotional agent, human's will trust it and like to interact with it more. Since these results occurred for both versions of prosody, we propose that these metrics are relatively robust to variations in timbre and prosody.

In our previous work, embedding prosody in robotic actions has been shown to increase perceived intelligence for individual robots. However in those studies the interactions were more social in nature [9], [8]. In our current experiment, where the robot was expected to perform a task (moving rings and placing them in a box) we propose that the successful performance by the robot was more influential on users' perception of its intelligence than external factors such as prosody.

Our initial hypothesis that willingness and confidence would be improved with prosody was not supported. In past work the effect of emotional prosody has not yet been used on individual robots for these two factors. It is not clear from this study whether prosody can influence these metrics which requires future study.

B. RQ 2: Entitativity and Prosody

Our results for research question 2 indicated that multiple voices did lower entitativity, increasing the perception of the group of robots as individual agents. This increase was achieved with only subtle variations, that could be easily achieved in real-time and scaled to many robots. We did not predict that having a single voice would increase entitativity however, as believed the gestures alone would appear as a group and single prosody would maintain this level. This reflected our original belief that entitativity would be relatively insusceptible to being increased amongst robots that already look and move in an identical manner. This finding has future implications for the possibility of audio design to not only reduce entitativity as per our original goal, but also the possibility of raising the level of entitativity.

C. RQ 3: Entitativity and HRI Metrics

A key finding in this study was the relation between entitativity and common HRI metrics. Our findings differ from those of related work on robots and groups [2]. This correlation between higher entitativity and each metric occurred across all groups independently, with gestures, single voice and multiple voice all showing the same relationship. We believe extensive future research should be undertaken to establish more completely the relationship between entitativity and groups of robots. We suggest that a possible explanation may be that with each robot performing the same task, participants may generally prefer interacting with the robot when perceived as a single agent, rather than having to engage with multiple agents. Multiple robots performing a similar task could give a perception that the robots are uniting towards a common goal. This would give participants a more positive impression that the robots are likeable and cooperative. This explanation would match Hamilton's studies on human groups that outsiders are more likely to engage with groups that have a higher entitativity [29].

D. Sound for Functionality

The majority of subjects text responses ranged from one to four sentences, and generally did not show much variation between groups. One stand out comment was that 8 participants from both prosody groups commented that they were not sure what the purpose of the sounds was. One participant noted: "I thought the singing was interesting but I don't see how that relates to the task success of the robot". Despite recognizing that the audio was not functional in the clip, this participant's ratings were well above the mean for each category, and we saw no reduction in ratings for any participant who noted there was no functional purpose. Nevertheless, we believe there is significant possibilities in

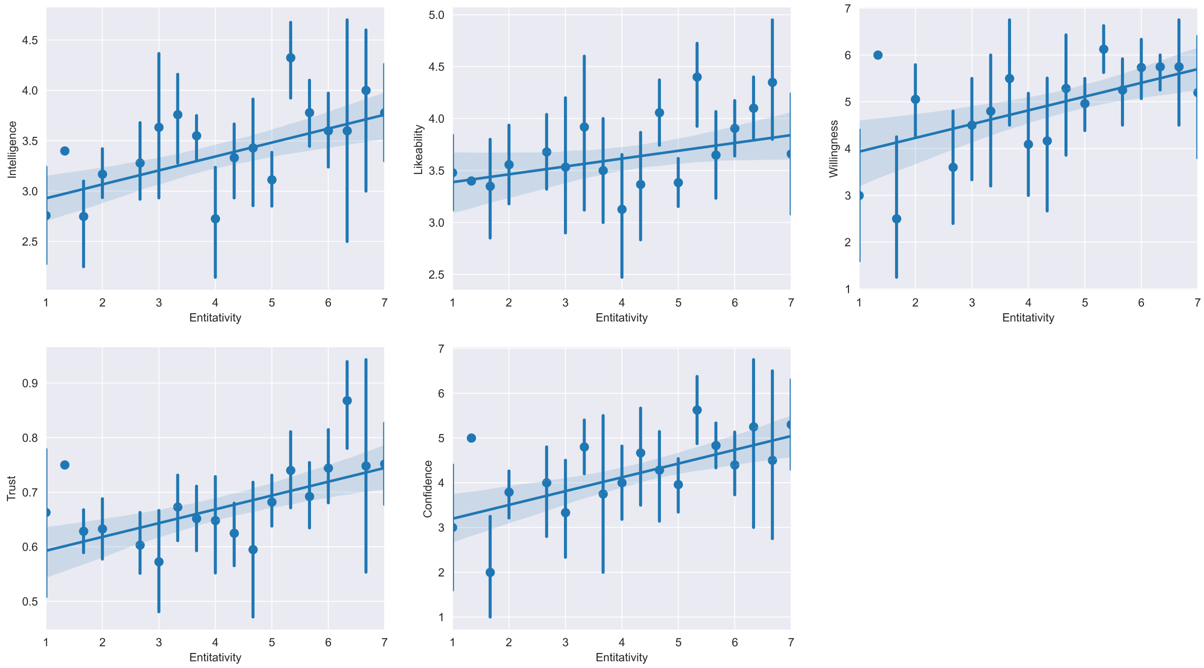

Fig. 6. Linear Regression comparing entitativity with HRI Metrics

considering the different applications of functional compared to non-functional or auxiliary sound and understanding how that impacts individual as well as groups of robots.

E. Limitations

This study was performed online using pre-recorded videos instead of live interaction or video watching in person. We believe that for this experiment this was an acceptable experimental design as ultimately our analysis focused on external viewing and analyzing a group of robots. Multiple past papers have shown no significant variation in results when a participant is watching a robot on video compared to in person [37], [38]. We also believe the use of MTurk and Prolific has significant advantages over in person studies, allowing us a far larger and more diverse participant pool than possible in person. It has also been shown that compared to university pools, MTurk participants are more careful [39]. When combined with our multiple point attention check we are confident that our results would be replicated in person.

We chose to use an industrial arm as they are commonly used in group manufacturing settings. In future studies we are interested in researching how the impact of prosody on robotic groups varies between platforms such as social or humanoid robots. Likewise, we only compared prosody to no audio, and in the future expect to compare different audio conditions.

Like many HRI applications this experiment only occurred over a small time frame and did not consider long-term implications of the system [40]. The use of emotional prosody has not yet been studied in long-term applications, but may have different use cases and would require additional changes in the implementation. One participant commented on the time scale, describing: “I really like it in the short term but I feel like I’d get tired of it if I had to listen all day long”. In the future we are interested in applying the system to longer form interactions in person and considering how prosody can be adjusted for use not just in a single session.

VI. Conclusion

In this paper we have shown with significant results that embedding emotional driven prosodic sound in robotic group actions improves likeability of and trust in the robots. We also showed that variations in prosody can lead to lower levels of entitativity, however a single voice can raise the level of perceived entitativity. Our results analyzing the correlation between entitativity and other HRI metrics suggest a number of directions for future research to understand the wider impact entitativity has on collaborative robots. More broadly, we believe this work shows the extensive possibility of audio in robotic systems to improve human-robot interaction.

References

[1] M. R. Fraune, S. Sherrin, S. Šabanović, and E. R. Smith, “Is human-robot interaction more competitive between groups than between individuals?” in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 104–113.

[2] M. R. Fraune, S. Šabanović, E. R. Smith, Y. Nishiwaki, and M. Okada, “Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots,” in *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*. IEEE, 2017, pp. 205–213.

[3] A. L. Blanchard, L. E. Caudill, and L. S. Walker, “Developing an entitativity measure and distinguishing it from antecedents and outcomes within online and face-to-face groups,” *Group Processes & Intergroup Relations*, vol. 23, no. 1, pp. 91–108, 2020.

[4] N. Dasgupta, M. R. Banaji, and R. P. Abelson, "Group entitativity and group perception: Associations between physical features and psychological judgment." *Journal of personality and social psychology*, vol. 77, no. 5, p. 991, 1999.

[5] A. Bera, T. Randhavane, E. Kubin, A. Wang, K. Gray, and D. Manocha, "The socially invisible robot navigation in the social world using robot entitativity," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4468–4475.

[6] M. Mateas, "Artificial intelligence today," M. J. Wooldridge and M. Veloso, Eds. Berlin, Heidelberg: Springer-Verlag, 1999, ch. An Oz-centric Review of Interactive Drama and Believable Agents, pp. 297–328. [Online]. Available: http://dl.acm.org/citation.cfm?id=1805750.1805762

[7] 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.

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

[9] R. Savery, L. Zahray, and G. Weinberg, "Emotional musical prosody for the enhancement of trust in robotic arm communication," in *Trust, Acceptance and Social Cues in Human-Robot Interaction, RO-MAN 2020*, 2020.

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

[11] 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.

[12] S. S. Kwak, Y. Kim, E. Kim, C. Shin, and K. Cho, "What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot," in *2013 IEEE RO-MAN*. IEEE, 2013, pp. 180–185.

[13] T. Ende, S. Haddadin, S. Parusel, T. Wüsthoff, M. Hassenzahl, and A. Albu-Schäffer, "A human-centered approach to robot gesture based communication within collaborative working processes," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2011, pp. 3367–3374.

[14] X. Dou, C.-F. Wu, K.-C. Lin, and T.-M. Tseng, "The effects of robot voice and gesture types on the perceived robot personalities," in *International Conference on Human-Computer Interaction*. Springer, 2019, pp. 299–309.

[15] J. Monceaux, J. Becker, C. Boudier, and A. Mazel, "Demonstration: first steps in emotional expression of the humanoid robot nao," in *Proceedings of the 2009 international conference on Multimodal interfaces*. ACM, 2009, pp. 235–236.

[16] I. Maurtua, A. Ibarguren, J. Kildal, L. Susperregi, and B. Sierra, "Human–robot collaboration in industrial applications: Safety, interaction and trust," *International Journal of Advanced Robotic Systems*, vol. 14, no. 4, p. 1729881417716010, 2017.

[17] T. Lorenz, A. Mörtl, and S. Hirche, "Movement synchronization fails during non-adaptive human-robot interaction," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2013, pp. 189–190.

[18] A. Mehrabian and S. R. Ferris, "Inference of attitudes from nonverbal communication in two channels." *Journal of consulting psychology*, vol. 31, no. 3, p. 248, 1967.

[19] A. Vysocky and P. Novak, "Human-robot collaboration in industry," *MM Science Journal*, vol. 9, no. 2, pp. 903–906, 2016.

[20] A. M. Abrams and A. M. Rosenthal-von der Pütten, "I–c–e framework: Concepts for group dynamics research in human-robot interaction," *International Journal of Social Robotics*, pp. 1–17, 2020.

[21] C. Breazeal and L. Aryananda, "Recognition of affective communicative intent in robot-directed speech," *Autonomous robots*, vol. 12, no. 1, pp. 83–104, 2002.

[22] J. Crumpton and C. L. Bethel, "Validation of vocal prosody modifications to communicate emotion in robot speech," in *2015 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 2015, pp. 39–46.

[23] R. Savery, L. Zahray, and G. Weinberg, "Before, between, and after: Enriching robot communication surrounding collaborative creative activities," *Frontiers in Robotics and AI*, vol. 8, p. 116, 2021. [Online]. Available: https://www.frontiersin.org/article/10.3389/frobt.2021.662355

[24] R. Savery, "Machine learning driven musical improvisation for mechanomorphic human-robot interaction," in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21 Companion. New York, NY, USA: Association for Computing Machinery, 2021, p. 559–561. [Online]. Available: https://doi.org/10.1145/3434074.3446351

[25] L. Fortunati, A. Esposito, and G. Lugano, "Introduction to the special issue "beyond industrial robotics: Social robots entering public and domestic spheres"," 2015.

[26] M. A. Hogg, D. K. Sherman, J. Dierselhuis, A. T. Maitner, and G. Moffitt, "Uncertainty, entitativity, and group identification," *Journal of experimental social psychology*, vol. 43, no. 1, pp. 135–142, 2007.

[27] E. Castano, V. Yzerbyt, and D. Bourguignon, "We are one and i like it: The impact of ingroup entitativity on ingroup identification," *European journal of social psychology*, vol. 33, no. 6, pp. 735–754, 2003.

[28] E. Castano, "On the advantages of reifying the ingroup," *The psychology of group perception: Perceived variability, entitativity, and essentialism*, pp. 381–400, 2004.

[29] D. L. Hamilton, S. J. Sherman, and L. Castelli, "A group by any other name—the role of entitativity in group perception," *European review of social psychology*, vol. 12, no. 1, pp. 139–166, 2002.

[30] M. R. Fraune, S. Sherrin, S. Sabanović, and E. R. Smith, "Rabble of robots effects: Number and type of robots modulates attitudes, emotions, and stereotypes," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, pp. 109–116.

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

[32] N. L. Robinson, T.-N. Hicks, G. Suddrey, and D. J. Kavanagh, "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)*. IEEE, 2020, pp. 272–277.

[33] M. Chita-Tegmark, T. Law, N. Rabb, and M. Scheutz, "Can you trust your trust measure?" in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 92–100. [Online]. Available: https://doi.org/10.1145/3434073.3444677

[34] K. Kurebayashi, L. Hoffman, C. S. Ryan, and A. Murayama, "Japanese and american perceptions of group entitativity and autonomy: A multilevel analysis," *Journal of Cross-Cultural Psychology*, vol. 43, no. 2, pp. 349–364, 2012.

[35] B. Lickel, D. L. Hamilton, G. Wieczorkowska, A. Lewis, S. J. Sherman, and A. N. Uhles, "Varieties of groups and the perception of group entitativity." *Journal of personality and social psychology*, vol. 78, no. 2, p. 223, 2000.

[36] R. Savery, L. Zahray, and G. Weinberg, "Prosodycvae: A conditional convolutional variational autoencoder for real-time emotional music prosody generation," in *2020 Joint Conference on AI Music Creativity*. CSMC + MUME, 2020.

[37] S. Woods, M. Walters, K. L. Koay, and K. Dautenhahn, "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.* IEEE, 2006, pp. 750–755.

[38] S. Woods, M. Walters, Kheng Lee Koay, and K. Dautenhahn, "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.

[39] D. J. Hauser and N. Schwarz, "Attentive turkers: Mturk participants perform better on online attention checks than do subject pool participants," *Behavior research methods*, vol. 48, no. 1, pp. 400–407, 2016.

[40] B. Irfan, A. Ramachandran, S. Spaulding, D. F. Glas, I. Leite, and K. L. Koay, "Personalization in long-term human-robot interaction," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019, pp. 685–686.