1

Towards Improving User Experience and Shared Task Performance with Mobile Robots through Parameterized Nonverbal State Sonification

Authors

Liam Roy, Monash University, Liam.Roy@Monash.edu
Richard Attfield, Monash University, Richard.Attfield@Monash.edu
Dana Kulić, Monash University, Dana.Kulic@Monash.edu
Elizabeth Croft, University of Victoria, ecroft@uvic.ca

Abstract

Given the correct context, nonverbal interaction can express high-level information with greater universality, efficiency, and appeal than spoken words. The focus of this work is to develop a simple nonverbal communication strategy for conveying high-level robot information to facilitate human-robot collaboration. We propose a low-dimensional parameterized communication model based on nonverbal sounds (NVS). The proposed model functions by modulating a fixed number of parameters of a base sound in an attempt to communicate distinguishable high-level robot states. A valence-arousal mapping is used to characterize the continuous axes of the proposed two-dimensional parameterized model. The developed communication model is validated using an online interactive survey designed to explore how well the model communicates high-level robot information, and how this communication modality affects the user's experience and shared task performance. Specifically, we investigated three parameters: participants' perceived understanding of the robot whilst observing an interaction video, their willingness to continue observing interactions with the robot, and their estimation of the suitability of the proposed communication model for the given context. The results of this study provide insight and direction concerning to the use of simplified NVS communication for human-robot collaboration. In addition, this work builds support for the development of a positive feedback loop through this modality, encompassing positive user experience, increased interest in subsequent interaction, and increased collaborative performance via familiarization.

1.1 Intro

As we continue to integrate collaborative machines within critical economic sectors including manufacturing, logistics, and healthcare, a growing number of users with diverse backgrounds will enter collaborative interactions with robots, increasing the need for seamless human-robot interaction (HRI). Nonverbal communication forms an essential component of human interactions and has accordingly been an important focus in the development of human-robot interactions [1]. Given the correct context, nonverbal interaction can express information with more universality, efficiency, and greater appeal than spoken words [2,3]. Nonverbal cues have been used for both active transmission of specific information (*explicit*) and passive conveyance of state information (*implicit*). The combination of these two modalities has been shown to improve the mental models that humans develop for collaborative robots, thereby improving understanding and trust [2, 4, 5].

Despite it's advantages, nonverbal communication presents a trade-off related to the challenge of expressing complex ideas and the ambiguity of interpretation by human collaborators. This has been shown to occur even with seemingly intuitive nonverbal communication methods [6]. These disadvantages can be mitigated via human-robot familiarization. Improving user experience (UX) is an effective method for incentivizing users to spend more time with a nonverbal social robot, providing the opportunity to learn the mappings between it's nonverbal cues and corresponding states [7]. Thus, a positive feedback loop between UX and the learnability of a robot's nonverbal mapping could be realized, as studies show users who have a better awareness of a robot's internal state and intent are more likely to characterize interactions with that robot as interesting and enjoyable [2, 8].

Nonverbal cues including body language [9–11], gestures [3, 12, 13], facial expressions [2, 14], lights [8], sounds [15, 16] and colour [17] have been widely implemented among collaborative machines deployed in real-world settings. In addition, the use of nonverbal communication in robotics has proven to be an effective method for improving user experience (UX) and shared task performance [4,8,12,16]. The potential of expressive sounds for robot communication remains a relatively under-explored research area, with a general focus on emphatic emotions [18], such as joy, anger, and disgust. Less attention has been given to task-focused communicative expressions. Numerous developed state communication models map a single sound to a single state with minimal or no ability to incrementally transition between or vary within communicative states [15,19]. The ability to incrementally adjust sound parameters would enable robots to achieve a greater degree of fluidity in social interactions, creating smoother and more enjoyable interactions.

The focus of this work is to develop a low-dimensional parameterized communication strategy based on NVS for conveying high-level robot information to facilitate human-robot collaboration. In robotics, using sound to represent robot states and information is referred to as state sonification [16, 20]. The proposed model functions by modulating a fixed number of parameters of a base sound in an attempt to communicate distinguishable high-level robot states. Our work builds on previous studies of this interaction modality [15,16,20] with the goal of reducing miscommunication and improving the user's mental model of a collaborative robot. In this work, we investigate the efficacy of using

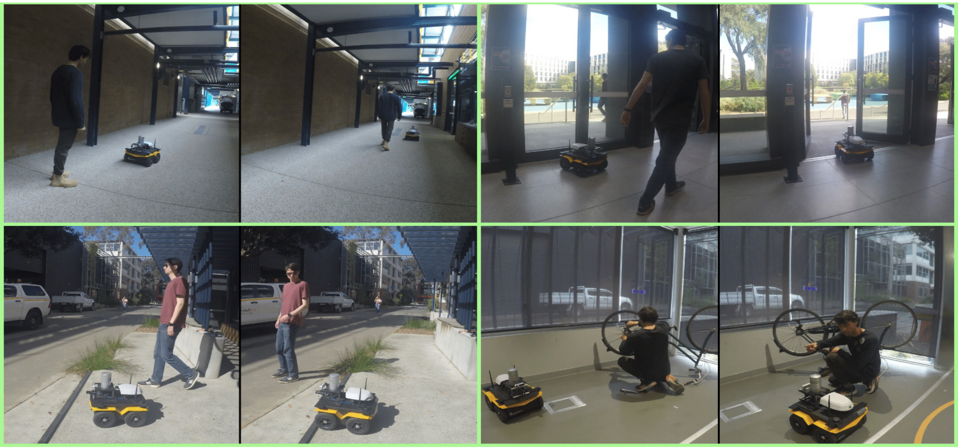


FIGURE 1.1: Key frames taken from video clips used in the online survey. (Top Left) Human initiating a leader-follower interaction with a robot. (Top Right) Robot attempting to communicate to a human that it needs assistance opening a door. (Bottom Left) Robot approaching a curb alongside a human. (Bottom Right) Human requesting assistance from robot while working on a bike.

parameterized state sonification to convey high-level robot information relevant to human-robot collaboration, and how this communication modality affects the user's experience and shared task performance.

The developed communication method is validated using an online user study. This study tested participant understanding of the proposed nonverbal communication strategy using audio clips and the suitability of the communication strategy for HRI scenarios using video clips. Fig.1.1 shows a set of sample images from the video set. The results of this study present both insight and direction concerning the use of simplified NVS communication for human-robot collaboration.

1.2 Related Work

Implicit vs Explicit Communication. Nonverbal communication consists largely of the implicit communication that can be observed in everyday interactions between people. While explicit communication, verbal or otherwise, is effective for direct and explicit communication, supplementing information implicitly can aid general understanding between collaborators. Using a collaborative task-based user study, Breazeal et al. showed that implicit nonverbal robot communication can improve a user's mental model of a robot's internal state, task efficiency, and error robustness [2]. Zinina et al. concluded that people greatly prefer interacting with a robot that utilizes implicit communication [5].

Implicit Communication in Non-Humanoid Mobile Robots. Implicit communication is typically modelled on human gestures and facial movements due to the ease with which people can pick up on these familiar expressions. Mutlu et al. found that users were able to better predict robot intent using cues provided by a humanoid robot's gaze [4]. Breazeal et al. found similar results using gaze combined with shrugging motions [2]. While less attention has been paid to implicit communication for non-humanoid mobile robots, previous studies have explored expressive lights [8] and sonification [20].

Nonverbal sounds for communication. Nonverbal sounds (NVS) have been used to communicate robot information in numerous ways, with inspiration often being drawn from the world of science fiction. Jee et al. used musical theory to analyse the sounds developed by Ben Burtt for the cinematic robots R2-D2 and Wall-E, from the films Star Wars and Wall-E, respectively [21]. Amongst others, this work concludes that the intonation of an NVS for robot communication should correlate to human speech. Similarly, Schmitz et al. used the concept of affect bursts, defined as "very brief, discrete, nonverbal expressions of affect in both face and voice as triggered by clearly identifiable events", to produce synthetic NVS to represent human emotional states [18]. Komatsu showed that altering a sound using a continuous parameter, in this case pitch change, can influence a human's perception of an artificial agent's interactive state by asking participants to match sounds to the states agreement, hesitation, and disagreement [22]. Luengo et al. proposed a model for NVS generation that splits sounds into indivisible sonic terms, or quasons [15]. An automated version of this model could combine different configurations of these terms based on situational context to represent different interaction states of a robot. To construct these quasons, three sound parameter categories (amplitude, frequency, and time) were identified and validated using an online questionnaire.

Sonification Mapping. In robotics, sonification is the process by which sounds are used to represent robot states and information. Sonification mapping is the process by which these states and sounds are related, and can take the form of emotion and action representations. Different sonification techniques include juxtaposing rhythmic vs. continuous sounds [16], the use of auditory icons-earcons [23, 24], and musical loops-based sonification [25]. Each technique has shown merit in accurately conveying specified actions, intent, or emotions. Recent publications have mapped music emotion [26], and more recently robot state sonification [16] using a 2D valence-arousal (VA) graph [27]. A rendition of this graph can be seen in Fig.1.2 and is further described in the *Methodology* section of this report.

1.3 Questions and Hypotheses

In this work, we investigated the efficacy of using parameterized state sonification to convey high-level robot information relevant to human-robot collaboration, and how this modality relates to UX and shared task performance. This work seeks to realize a positive feedback loop present between a user's interaction enjoyment and the learnability of a robot's nonverbal mapping. Previous studies have affirmed the use of nonverbal modalities to facilitate enjoyable interactions with social robots [2,8,12,28]. Informed by these prior works, we formulated the following research questions (Q1, Q2) and hypotheses (H1,H2):

Q1 How effective is a low-dimensional approach to parameterized state sonification for conveying high-level robot information relevant to human-robot collaboration?

Q2 How does user familiarization with this nonverbal communication strategy correlate to user experience (UX)?

H1 Linearly modulating two parameters of a sound will be a sufficient sonification strategy for communicating distinguishable high-level robot information relevant to human-robot collaboration.

H2 An appropriate sonification strategy will create a positive feedback loop, encompassing positive user experience, increased interest in subsequent interaction, and increased collaborative performance via familiarization.

The remaining sections of this chapter are structured as follows. The *methodology* and *user study* sections detail the approach we followed to answer the stated research questions and test our associated hypotheses. We then summarize our findings in the *results & analyses* section, followed by a *discussion* of these results. Finally, we close with final remarks and offer our thoughts on interesting avenues for relevant *future work*.

1.4 Methodology

Collaborative Robot States. In a related study, Baraka et al. [8] modelled three high-level robot states (on-task, stuck, requesting help) which most accurately represented their robot throughout their interaction scenario using onboard lights. Extrapolating from [8] and other studies [4,6,20] which focused on communicating robot states relevant to human-robot collaboration, we formulated five high-level robot states relevant to human-robot collaboration: *idling*, *progressing*, *successful*, *unsuccessful* and *requires attention*. Each of these states is briefly described in Table 1.1 and visualized as a continuous region along a two-dimensional valence-arousal (VA) graph in Fig.1.2.

The state *idling* is defined as a robot at rest, awaiting human-initiated interaction. This state is analogous to the standby mode common to personal electronics such as computers, televisions and mobile phones. In this state, the robot is not communicating auditory information. The state occupies the lower region of the VA graph, Fig.1.2 where arousal=0. Above *idling*, the state *progressing* occupies the central region of the graph. The state *progressing* overlaps with most other non-mutually exclusive states, as a robot can be actively processing a task with ranging degrees of confidence (correlated with valence) and urgency (correlated with arousal). On the right-hand region of the graph, the state *successful* is reached when the robot completes an assigned task. The overlapping region between *successful* and *progressing* refers to an on-task robot that is confidently progressing through an assigned task, whereas the right-most region refers to a robot which has successfully completed an assigned task.

The left-hand region of the graph *unsuccessful* is reached when a robot is unable to complete an assigned task. The overlapping region between *unsuccessful* and *progressing* refers to an on-task robot that is confused or struggling with an assigned task, whereas the left-most region refers to a robot which has failed or is unable to complete an assigned task entirely. Finally, the upper region of the graph refers to the state *requires attention*. In this state, a robot is attempting to get a human's attention and initiate an interaction.

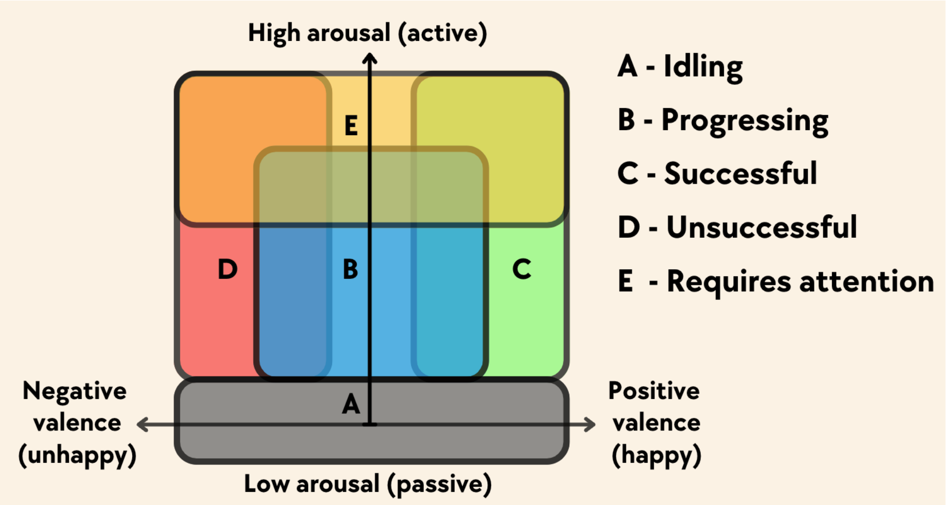


FIGURE 1.2: Visualization of Five High-Level Robot States on a Valence-Arousal Graph.

Collab. State	Description
Idling	At rest, awaiting human-initiated interaction
Progressing	On-track, actively processing an assigned task
Successful	Confident or able to complete an assigned task
Unsuccessful	Struggling or unable to complete an assigned task
Req. Attention	Attempting robot-initiated interaction

TABLE 1.1: Summary of Five Formulated Robot States Relevant to Human-Robot Collaboration.

Similar to *progressing*, this state overlaps other states, as a robot may be actively seeking attention for positive or negative reasons.

Sonification Model Design. To develop our nonverbal state sonification model, we investigated sonification techniques proven to facilitate effective and enjoyable interaction with humanoid and social robots [2, 20]. This investigation aimed to build a sonification model for a non-humanoid mobile robot based on sonification techniques which translate intuitively from humanoid and social robotics. We explored a sonification strategy in which we modulated a fixed number of parameters of a base sound in an attempt to communicate distinguishable high-level robot states. We focused on a low-dimensional model, strictly adjusting two parameters while keeping all remaining parameters constant, to reduce the number of variables and simplify the model validation process. By restricting the number of communication variables, we sought to reduce the required complexity of translation between the robot's state information and it's communication. Similar to [16], a 2D valence-arousal mapping [27] was used to characterize the continuous axes targeted by this 2D sound parameterization, as shown in Fig.1.2. Using a model initially developed for human emotional communication [29] presented the opportunity to explore different sound parameters that could map effectively to these axes, such as frequency to valence or volume to arousal.

To reduce the number of variables and simplify the model validation process, we structured our sonification model around one neutral sound set on a loop. This sound was selected based on it's neutral characteristics, namely, it's short duration and consistent pitch and volume. We situated this neutral sound at the origin of our VA graph. The sample we used was a 4-second synth sound file named *Infuction_ F#.aif* from the Ableton Live 10 sample library, shown in Fig.1.3[1].

We experimented with mapping different sound parameters to each axis of our 2D VA graph. Parameters that were explored for the valence axis included the sound's relative harmonic key, pitch, and amplitude of pitch change within a single loop. For each experimental mapping, we used the digital audio workspace Ableton Live 10 to modulate the neutral sound situated at the origin. This output sound was used to gauge how the modulation of each candidate parameter correlated to perceived changes in the neutral

[1] Sonification library: https://tinyurl.com/sonificationlibV1

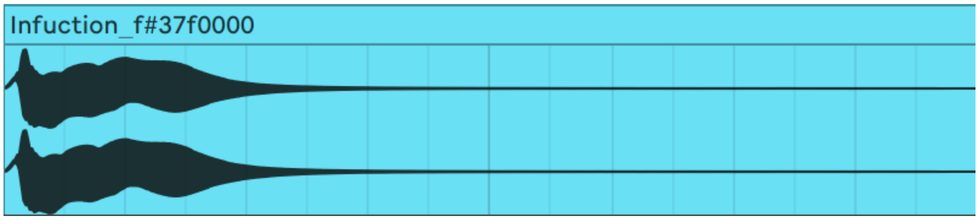


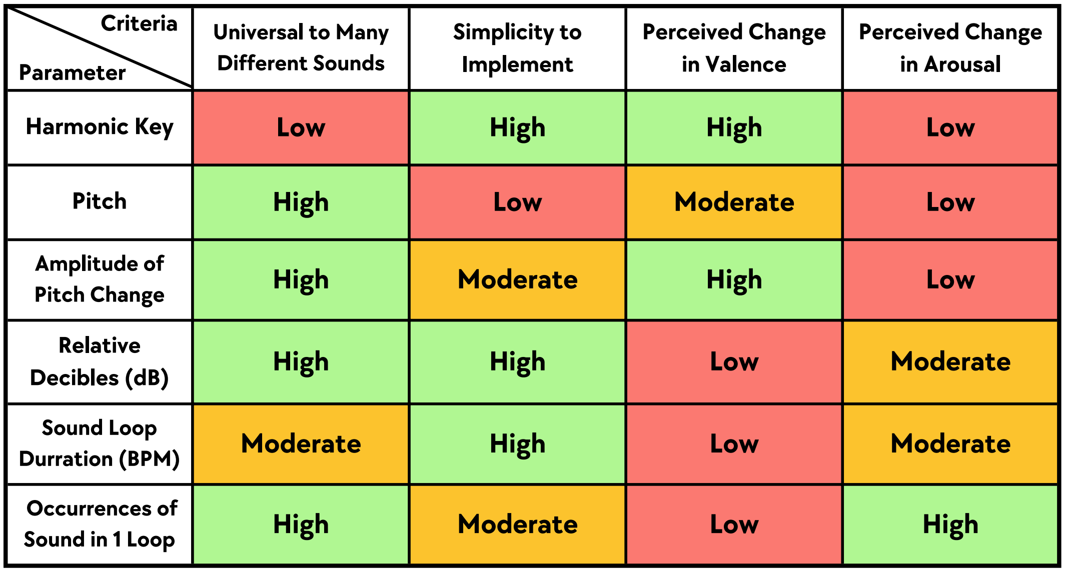
FIGURE 1.3: Visualization of the Neutral Sound File Named *Infuction_ F#.aif*. The dark-shaded region represents the audio file's sound wave. In this representation, the x-axis is the time domain, while the y-axis is the magnitude of the sound wave. Translation along the y-axis represents changes in the wave's frequency. The thickness of the wave is it's volume. The sound wave is shown twice as the audio file is stereo, meaning the file has both a unique sound wave for left and right channels in a 2-channel audio system. In this case, both left and right sound waves are identical.

sound's valence. The same process was done for the arousal axis, exploring parameters such as relative decibels (dB), the sound loop's duration (BPM), and the number of occurrences of the neutral sound within a single loop. Fig.1.4 shows a decision matrix used to identify two parameters from an initial list of six to use for the 2D sonification mapping. The criteria used to identify a suitable pair were the perceived changes to valence and arousal, the potential of the parameterization to be applied to any arbitrary base sound with previously stated neutral characteristics, and the simplicity of implementation. The performance measures shown in Fig.1.4 were estimated following an initial analysis of the considered parameters.

From the decision matrix shown in Fig.1.4, the parameter selected for the valence axis of our 2D mapping was the amplitude of pitch change within a single loop. Changes in this parameter were estimated to have a high impact on the perceived valence, with a lower effect on arousal. This parameterization was also deemed relatively simple to implement, and easy to apply to different base sounds. The final parameter selected for the arousal axis of our 2D mapping was the number of occurrences of the neutral sound within a single loop. This parameter was perceived to have a corresponding impact on arousal without affecting valence and was also estimated to be comparably simple to implement and apply to different base sounds. As well as being individually appropriate, these two parameters were estimated to be a suitable complement to one another.

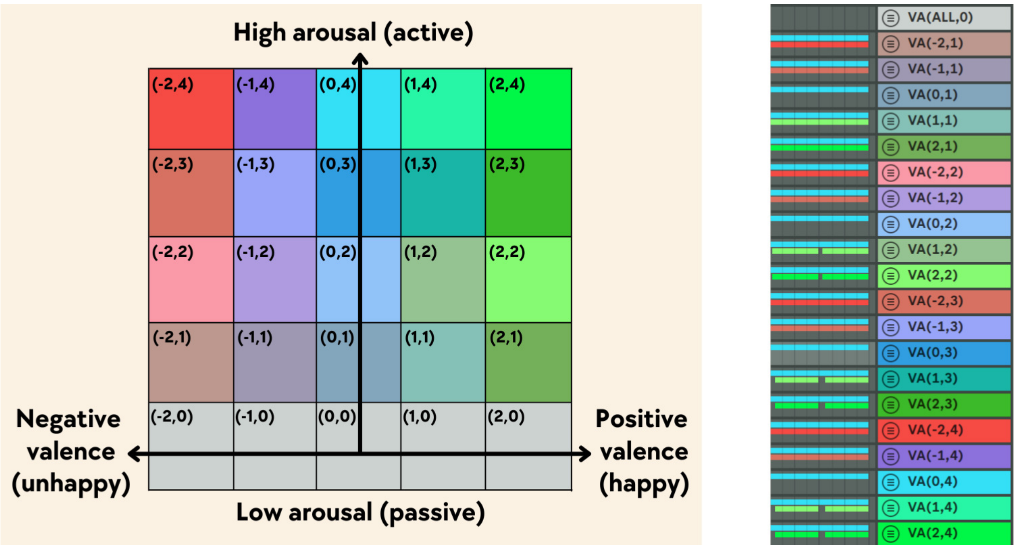
Using the two selected parameters, the VA graph shown in Fig.1.5 was discretized into 25 regions with valance ranging from [-2, 2] and arousal ranging from [0, 4]. At the region (0,1), the neutral sound is played once within the communication loop, with no change in pitch through the loop. As shown in Fig.1.6, increases and decreases along the discretized valence axis represent weighted positive or negative pitch change throughout a single communication loop. As shown in Fig.1.7, increases and decreases along the discretized arousal axis represent adding or removing repetitions of the neutral sound within a single communication loop. No sound is produced in the grey regions in Fig.1.5 where arousal is zero.

The discretized regions shown on the VA graph in Fig.1.5 directly map to the selected high-level states represented as continuous regions in Fig.1.2. While regions at the center of the graph (-1,1) through (1,3) map to the state *progressing*, regions along the top of the graph (-2,3) through (2,4) map to *requires attention.* Similarly, regions (1,1) through (2,4) along the right-hand side map to *successful*, and regions (-2,1) through (-1,4) along the left-hand side map to *unsuccessful.* As previously discussed, *Idling* is represented by the greyed-out regions along the bottom of the graph (-2,0) to (2,0) where no sound is produced.



Criteria / Parameter	Universal to Many Different Sounds	Simplicity to Implement	Perceived Change in Valence	Perceived Change in Arousal
Harmonic Key	Low	High	High	Low
Pitch	High	Low	Moderate	Low
Amplitude of Pitch Change	High	Moderate	High	Low
Relative Decibles (dB)	High	High	Low	Moderate
Sound Loop Durration (BPM)	Moderate	High	Low	Moderate
Occurrences of Sound in 1 Loop	High	Moderate	Low	High

FIGURE 1.4: Decision matrix used to identify appropriate parameters for the proposed sonification model.



(-2,4)	(-1,4)	(0,4)	(1,4)	(2,4)
(-2,3)	(-1,3)	(0,3)	(1,3)	(2,3)
(-2,2)	(-1,2)	(0,2)	(1,2)	(2,2)
(-2,1)	(-1,1)	(0,1)	(1,1)	(2,1)
(-2,0)	(-1,0)	(0,0)	(1,0)	(2,0)

VA(ALL,0)
VA(-2,1)
VA(-1,1)
VA(0,1)
VA(1,1)
VA(2,1)
VA(-2,2)
VA(-1,2)
VA(0,2)
VA(1,2)
VA(2,2)
VA(-2,3)
VA(-1,3)
VA(0,3)
VA(1,3)
VA(2,3)
VA(-2,4)
VA(-1,4)
VA(0,4)
VA(1,4)
VA(2,4)

FIGURE 1.5: (Left) Visualization of 25 discretized regions on a valence-arousal graph. (Right) Sound files in Ableton Live 10 corresponding to each discretized region on the valence-arousal graph.

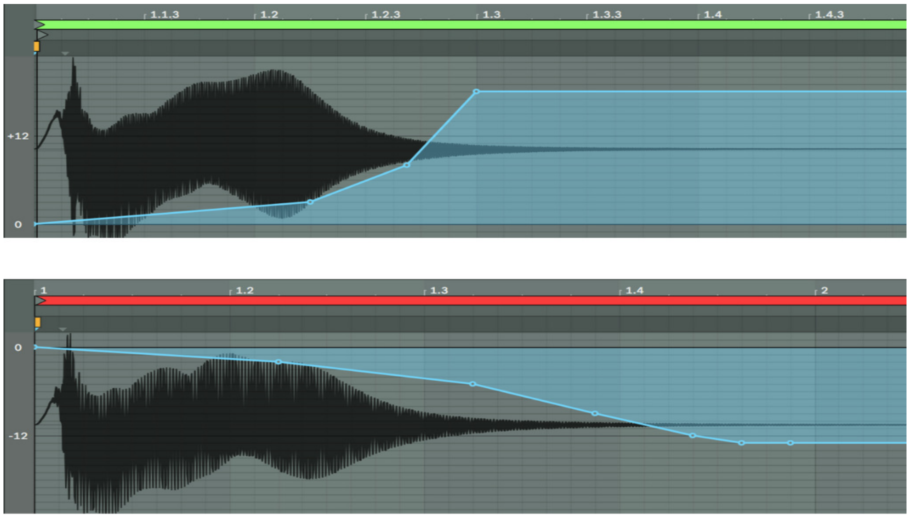


FIGURE 1.6: (Top) Visualization of the positive pitch modulation applied to the neutral sound wave to produce a valence=1 sound. (Bottom) Visualization of the negative pitch modulation applied to the neutral sound wave to produce a valence=-1 sound. Note the axes of the left of both figures, representing the amplitude of pitch change for both pitch modulation examples.

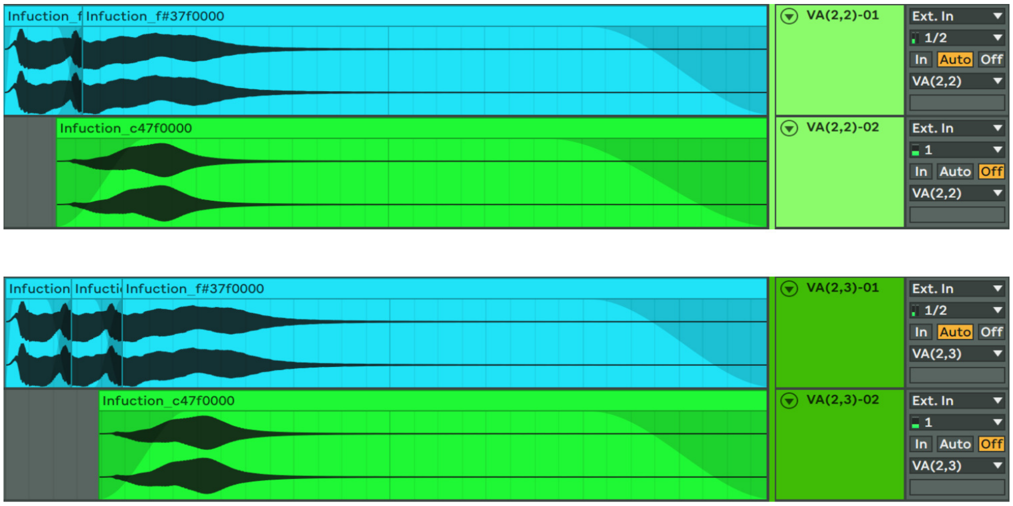


FIGURE 1.7: (Top) Visualization of the sound profile which represents the discretized region VA=(2,2). Arousal=2 is represented by two repetitions of the blue-shaded neutral sound wave. (Bottom) Visualization of the sound profile which represents the discretized region VA=(2,3). Arousal=3 is represented by three repetitions of the blue-shaded neutral sound wave. (Both) The green shaded sound wave represents the positive valence=2 sound blended with the repetitions of the neutral sound.

1.5 User Study

Structure of Survey. To validate the developed sonification model, we carried out an online observational study using the platform Qualtrics[2]. This user study was reviewed and approved by the Monash University Human Research Ethics Committee (MUHREC) with project ID 35703. This survey contained a series of audio samples from our sonification model, videos of HRI scenarios shot using a GoPro camera[3], and associated interactive questions. This study was formatted as an observational survey to reduce possible stimulus variation between participants. As outlined by [30] and demonstrated by [31–33], screen-based methods are an effective way to fix the exact stimulus that each participant experiences. To this end, participants of our study were asked to listen to the same audio samples and observe the same videos. This study was conducted as an online survey, such to facilitate the recruitment of a large, diverse pool of participants. This also followed in the mould of previous similar studies, such as [15].

Before commencing the main sections of the survey, participants were asked to complete a preliminary data-collection consent form, along with a set of pre-survey questions. These questions were used to collect demographic data along with self-scored experience and enthusiasm levels related to working with collaborative robots. Upon completing this preliminary section, users were brought to an introductory page outlining the concepts of valance and arousal and their relation to robot states. This introduction was followed by six questions in which users were presented with a blank VA graph discretized into 25 regions as shown in Fig.1.5. For each of these six questions, users were asked to listen to a sound selected from our sonification library and guess which region this sound represented on the graph. The six sounds were chosen to give a distributed representation of the VA graph. A visualization of this selection process is shown in Fig.1.8.

The next section of our survey was similar to the previous one, with the addition of videos. For each question, users were presented with a video in which a Jackal mobile

[2] Qualtrics User Study: https://tinyurl.com/qualtricsstudyV1

[3] YouTube playlist of all survey videos: https://tinyurl.com/sonificationvideos

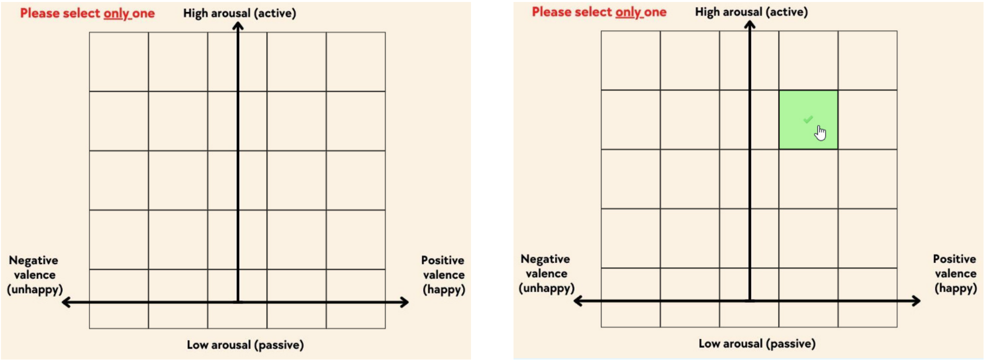


FIGURE 1.8: (Left) VA graph with 25 discretized selection regions presented to users. (Right) VA Graph after users have made their selection.

robot was communicating with a human using the designed sonification model. Each video captured a unique HRI scenario with a unique outcome. The video was paused at a key point midway through the interaction scenario, at which point users were asked which region the sounds emitted from the Jackal represented on the discretized VA graph. Keyframes from these video clips are shown in Fig.1.1. In addition to this selection, users were asked to answer a multiple-choice question regarding which state they perceived the robot to be in. The options for this question consisted of each state outlined in Table 1.1 with the additional option *not sure*.

Unlike the previous section with a fixed number of questions (six), participants selected the number of video questions they wanted to complete. Upon completing an initial mandatory set of five video questions, users were presented with an option to view another video or proceed to the final video question. The number of video questions that users could complete within this section was capped at 15. Allowing users to select whether they would like to watch another video was used to gauge their interest in the robot interactions.

Following this set of video questions, users were presented with a final video question. This video was identical to the first video users watched in the previous section. The repeat video was used to gauge how contextual familiarization with the sonification model affected their ability to learn the sonification model and thus improve their robot state estimation accuracy. In addition, the variability in the number of video questions completed in the previous section presented an indication of the effects of extended familiarization with the sonification model.

Upon completing the repeated video question, users were presented with another section in which they were asked to listen to sounds without videos and guess which region these sounds represented on the discretized VA graph. Unique sounds from those presented at the beginning of the study were used in this section. Finally, participants were presented with a set of post-survey questions tailored from the BUZZ scale [34]. These questions probed for feedback on the audio communication model and user study overall.

Participants. This online survey was distributed as a single link over multiple Monash University social pages, along with several external networking platforms not affiliated with the university. In all, our study received 37 complete responses. An analysis of our collected demographics data revealed an overall age spread of 18-60+, with 43% of respondents in their 20s. In addition, 65% of respondents self-reported having little (2) to no (1) experience working with collaborative robots on a 1-5 Likert scale. To our surprise, there was a wide range of professions among participants including data analysts, government workers, teachers, and students.

1.6 Results and Analysis

Sonification testing. To test the intuitiveness and learnability of the sonification mapping, the errors in the participants' predictions of a sound's position on the VA graph were recorded. For simplicity, the error was recorded as the distance in grid squares of the prediction from the true position on the discretized VA grid. Fig.1.9 shows the trends of errors across the question set. There is a trend towards decreasing errors across the question set, although with a low statistical significance.

Further analysis of the errors in VA predictions is shown in Fig.1.10. These distributions show that the errors in arousal prediction were higher than those in valence, although with a large overlap. A 15% decrease in averaged overall participant error was observed in the second batch of sounds as opposed to the first, indicating a degree of learnability in the designed sonification method.

Interaction scenario testing. For each interaction video, participants were able to select up to two robot states that they believed to apply to the given scenario. The low arousal state, *Idle*, was mapped to a non-communicative state in which no sound was emitted from the robot. Thus, this state was not analyzed in the set of interaction videos. Fig.1.11 shows the percentage of participants who were able to correctly identify each high-level robot state, averaged over the question set. States *successful* and *requires attention* were both identified correctly by over 70% of participants across the question set, while *progressing* and *unsuccessful* were both identified by roughly 50% of participants.

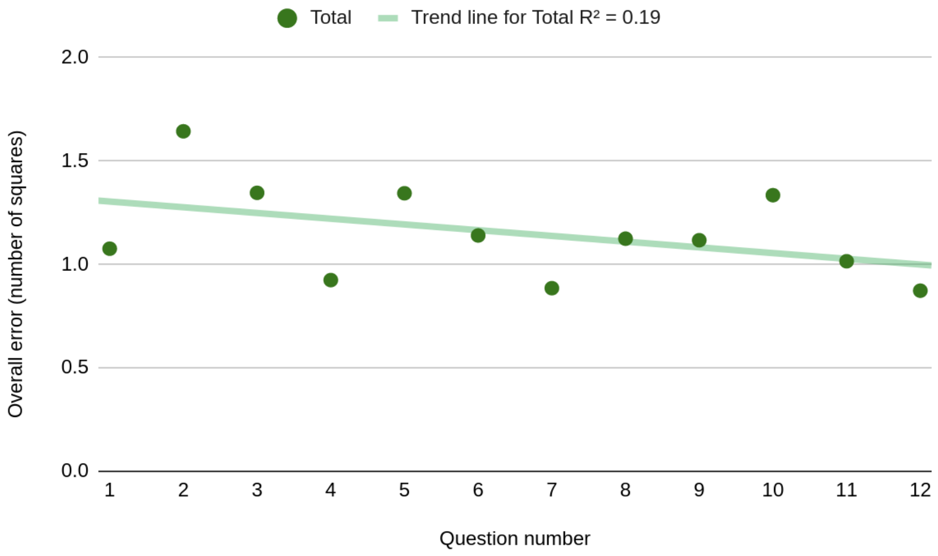


FIGURE 1.9: Error in VA predictions across sound question set

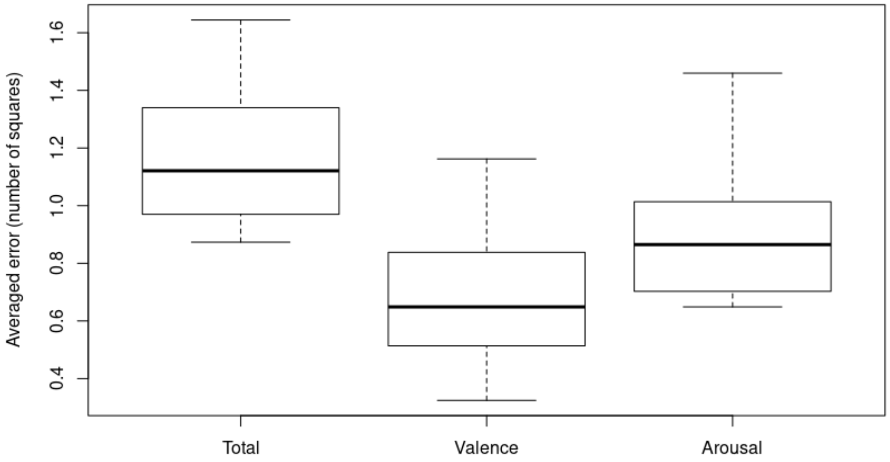


FIGURE 1.10: Error distributions across sound question set

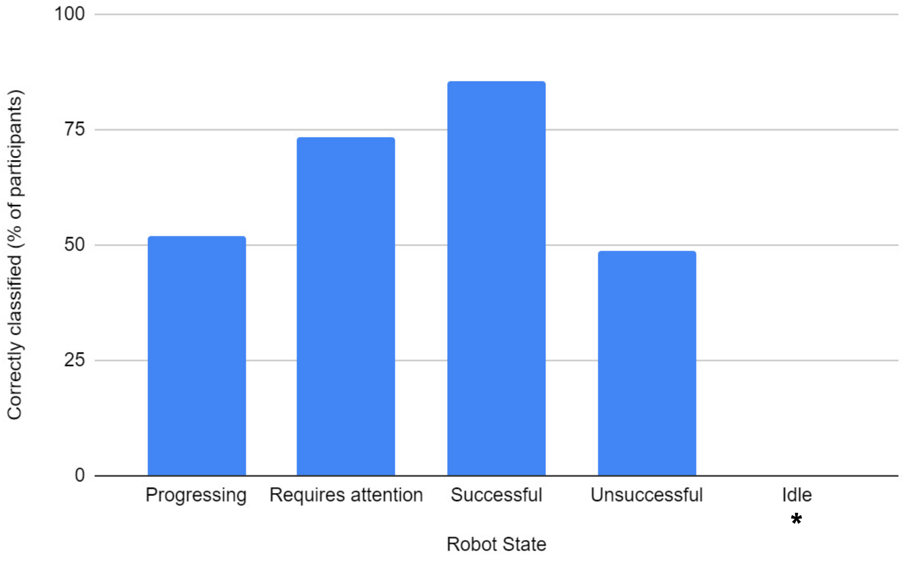
As participants were able to select multiple states for each scenario, further analysis is provided by examining the correct state selections in relation to the total selections made by participants. Fig.1.12 shows the participants' selection accuracies for each robot state using the *correct* and *incorrect* scores. These scores were calculated using the correct and incorrect selections as percentages of the total selections for each question (i) averaged over the question set (m) as shown in the following equations:

$$correct\_score = \frac{1}{m}\sum_{i=1}^{m}\frac{correct\_selections_i}{total\_selections_i} * 100\%$$

$$incorrect\_score = \frac{1}{m}\sum_{i=1}^{m}\frac{incorrect\_selections_i}{total\_selections_i} * 100\%$$

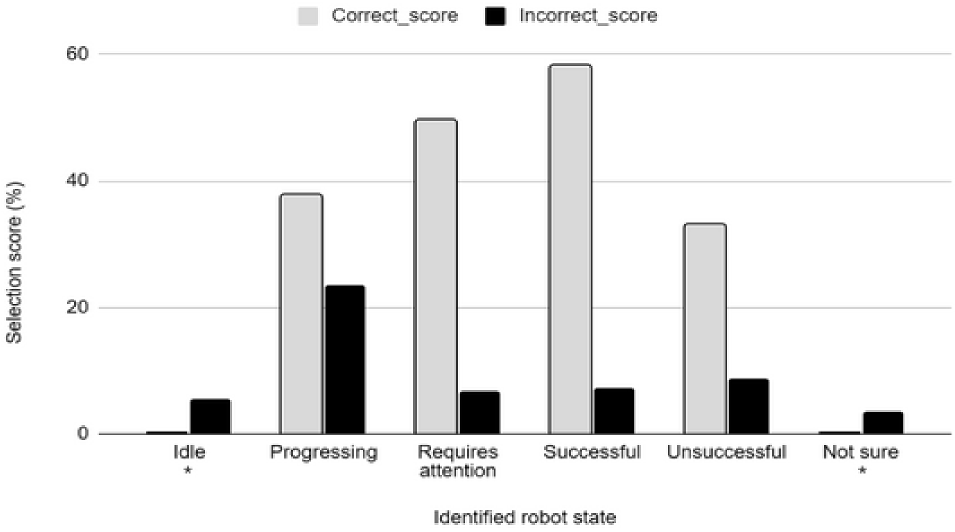
Fig.1.12 shows some strengths and weaknesses of the developed sonification method. *Successful* was found to be easier to convey than *unsuccessful*, with a 25% higher average classification accuracy. Similarly, the high arousal state *requires attention* was correctly classified more than the lower arousal state, *progressing*, with a 12% margin. Overall, the incorrect selection scores of all classes were relatively low, with the exception of the *progressing* class.

No strong trend was observed in the classification accuracy across the question set, as the specific situational context was found to have a dominant effect. However, by analyzing the change in performance for the repeat question, a positive effect of familiarisation with



***** No interaction scenarios were developed for this state

FIGURE 1.11: Participants' prediction accuracies for each robot state



* No interaction scenarios were developed for these state options

FIGURE 1.12: *Correct* and *incorrect* scores for each robot state

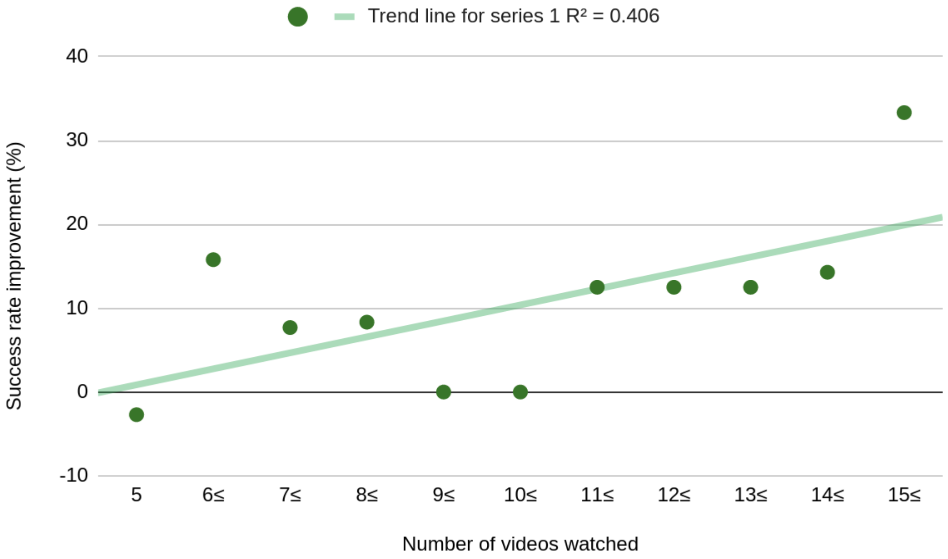


FIGURE 1.13: Effect of repeated interaction on duplicate scenario improvement

the Jackal's communication model may be substantiated. In general, there was little difference seen between the first question and the duplicate final question. However, as shown in Fig.1.13, a correlation between the number of videos a participant elected to watch and the classification accuracy improvement on the final repeated question was observed.

Fig.1.14 shows the effect of the participants' self-recorded feelings towards the communication model on the number of videos they chose to watch. In general, there is an observed relationship between fondness for the communication model and a desire for repeated interaction via the survey videos. Paired with Fig.1.13, this points towards a positive relationship between user experience and robot understanding. However, this cannot be conclusively attributed to the communication method, as fondness for the robot's communication cannot be concretely separated from fondness for the robot itself.

When analyzing the valence-arousal predictions made during the video section of the user study, a larger spread of errors was observed, as shown in Fig.1.15. There was no observable trend in these errors across the video set. This emphasizes the large role that situational context played in the participants' understanding of the robot during real-world interactions.

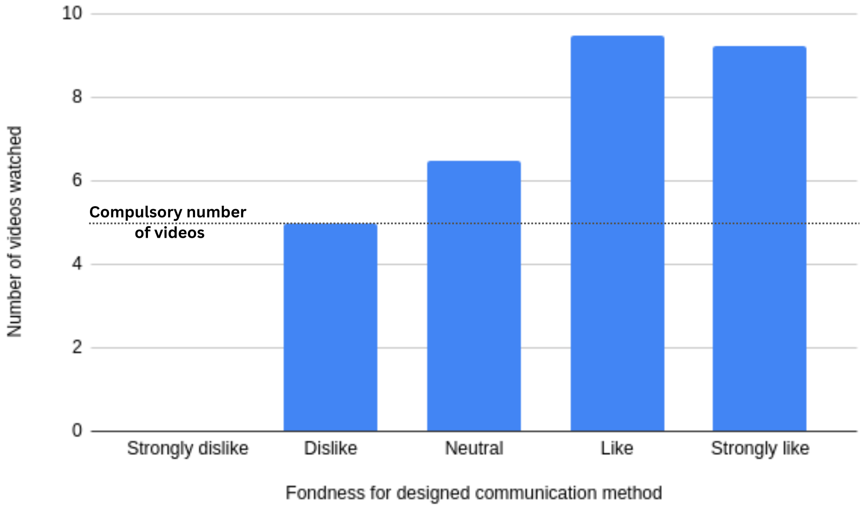


FIGURE 1.14: Average number of videos watched according to self-recorded feeling towards the communication model

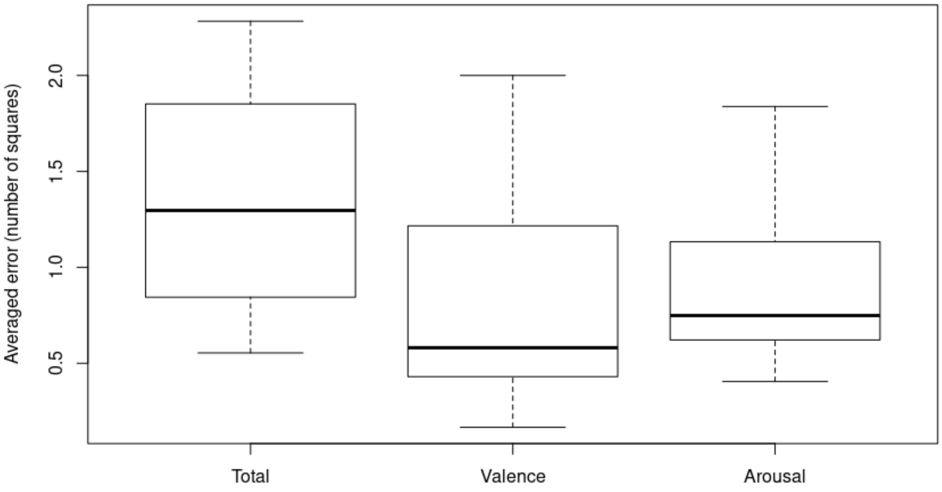


FIGURE 1.15: Distribution of errors in valence-arousal prediction of video section

1.7 Discussion

The data presented in Fig.1.11 and Fig.1.12 show that a low-dimensional continuous sonification mapping can be used to communicate high-level robot state information, supporting **H1**. However, certain states were understood by the participants with greater accuracy than others, indicating the need for further refinement in the designed communication model. Due to the use of lower frequencies, low-valence sounds were more difficult to differentiate than high-valence sounds on standard audio devices. This may explain why the *successful* interactions were classified with a higher degree of accuracy when compared against *unsuccessful*. Additionally, multiple participants reported that the *neutral* position on the VA graph sounded too high in arousal. This could explain why the high arousal state *requires attention* was correctly identified at a higher rate than the lower arousal state *progressing*, as participants regularly mistook passive communication for more active. The relatively high *incorrect* score of the *progressing* class may be attributable to it's position relative to the other states, as participants unsure of their predictions were perhaps likely to also select this more neutral state.

Overall, the majority of participants rated the communication as "Easy" or "Very easy" to understand, however, over a fifth rated the communication as "Difficult" or "Very difficult", highlighting that the proposed strategy is not intuitive to all users. This breakdown is shown in Fig.1.16. The participants' feelings towards the communication model were determined using both a Likert scale question, shown in Fig.1.17, and a set of open-ended questions. Both mediums showed dominantly favourable reactions with a few negative responses. The correlation between likability and number of videos watched, shown in Fig.1.14, points to a potential relationship between user experience and interest in repeated interaction. Viewed in conjunction with Fig.1.13, the realized positive feedback loop between user experience, repeated interaction, and consequently improved robot understanding begins to support **H2**. A potential advantage key to this positive feedback loop is the improved user enjoyment resulting from a non-traditional, interesting communication form. Similar to **H1**, properly validating **H2** would require a comparison analysis between alternate nonverbal and verbal communication strategies.

With respect to the validation parameters, we found that participants were able to understand certain high-level robot states within different interaction scenarios. The relative difficulty observed in communicating certain states is believed to be due to features of the sonification model. From these findings, we believe that the model could be improved by improving sound clarity at low valence states and tweaking the parameterization techniques to better differentiate between high and low arousal states.

Regarding the perceived suitability of the proposed communications strategy, Fig.1.16 and Fig.1.17 indicate that this communication strategy is both favourable and intuitive to most participants, acknowledging some discrepancy among participants and the small size of the study pool. This discrepancy is expected, as ambiguity is unavoidable in all non-explicit communication and may result in a negative reaction from some users.

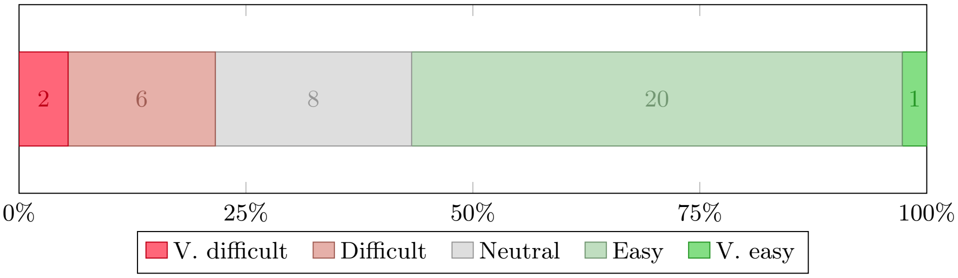


FIGURE 1.16: Participants' ease of understanding communication(absolute participant numbers shown on graph)

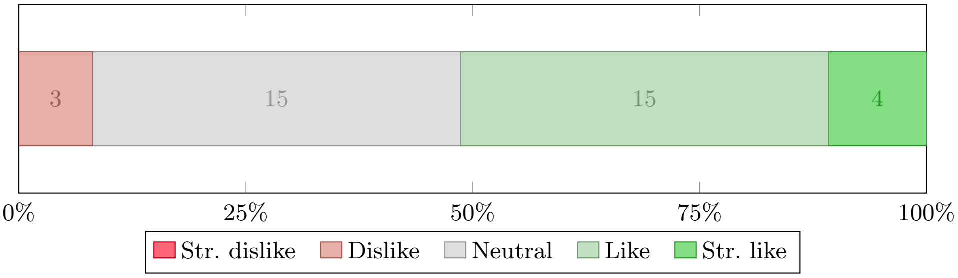


FIGURE 1.17: Participants' reactions to communication method (absolute participant numbers shown on graph)

1.8 Conclusions and Future Works

Reflecting on our research questions:

Q1 How effective is a low-dimensional approach to parameterized state sonification for conveying high-level robot information relevant to human-robot collaboration?

Q2 How does user familiarization with this nonverbal communication strategy correlate to user experience (UX) and shared task performance?

In our study, the proposed NVS model was able to effectively communicate the outlined set of robot states, with a degree of error due to the simplicity of the proposed 2D mapping. The findings also show variance between the different states; communicating the states *successful* and *requires attention* were less error-prone than *unsuccessful* and *progressing*. This finding presents an opportunity to experiment with alternate approaches for expressing mid-ranged arousal and negative valence. In the interest of making the sounds within our model more distinguishable, we plan to explore the use of different base sounds, blending sounds at the extremities of our VA mapping with auditory icons-earcons [20, 23], reducing the number of sounds emitted while *progressing*, correlating volume with arousal, rapidly oscillating the pitch of negative valence sounds rather than simply bending the pitch in the negative direction, and increasing our sonification model's dimensions of parameterization. Ultimately, we believe adding more complexity to our communication model while retaining structured parameterization would be an effective way to generate more meaningful and appealing sounds while further reducing human-robot miscommunication. To aid in characterizing the modulation of an additional sound parameter, we plan to explore the extension of a 2D valance-arousal (VA) mapping to a 3D valance-arousal-stance (VAS) mapping. This extension has proven effective for conveying robot emotion with a higher degree of complexity [35, 36]. An interesting avenue of future work would be to explore how the use of this higher-dimensional mapping might be useful for conveying high-level robot information relevant to human-robot collaboration.

Bibliography

[1] Alessandro Marin Vargas, Lorenzo Cominelli, Felice Dell'Orletta, and Enzo Pasquale Scilingo. Verbal Communication in Robotics: A Study on Salient Terms, Research Fields and Trends in the Last Decades Based on a Computational Linguistic Analysis, 2 2021.

[2] Cynthia Breazeal, Cory D. Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pages 708–713, 2005.

[3] Henny Admoni, Thomas Weng, Bradley Hayes, and Brian Scassellati. Robot nonverbal behavior improves task performance in difficult collaborations. Technical report, 2016.

[4] Bilge Mutlu, Fumitaka Yamaoka, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Nonverbal leakage in robots. page 69. Association for Computing Machinery (ACM), 2009.

[5] Anna Zinina, Liudmila Zaidelman, Nikita Arinkin, and Artemiy Kotov. Non-verbal behavior of the robot companion: A contribution to the likeability. *Procedia Computer Science*, 169(2019):800–806, 2020.

[6] Rolando Fernandez, Nathan John, Sean Kirmani, Justin Hart, Jivko Sinapov, and Peter Stone. Passive Demonstrations of Light-Based Robot Signals for Improved Human Interpretability. *RO-MAN 2018 - 27th IEEE International Symposium on Robot and Human Interactive Communication*, pages 234–239, 2018.

[7] Richard Savery, Lisa Zahray, and Gil Weinberg. Emotional musical prosody for the enhancement of trust: Audio design for robotic arm communication. *Paladyn*, 12(1):454–467, 2021.

[8] Kim Baraka, Stephanie Rosenthal, and Manuela Veloso. Enhancing human understanding of a mobile robot's state and actions using expressive lights. *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016*, pages 652–657, 2016.

[9] Heather Knight and Reid Simmons. Laban head-motions convey robot state: A call for robot body language. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, volume 2016-June, pages 2881–2888. IEEE, 5 2016.

[10] Arjun Sripathy, Andreea Bobu, Zhongyu Li, Koushil Sreenath, Daniel S. Brown, and Anca D. Dragan. Teaching Robots to Span the Space of Functional Expressive Motion. 2022.

[11] Martin Inderbitzin, Aleksander Valjamae, Jose Maria Blanco Calvo, Paul F.M.J. Verschure, and Ulysses Bernardet. Expression of emotional states during locomotion based on canonical parameters. *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, FG 2011*, pages 809–814, 2011.

[12] Anca D. Dragan, Shira Bauman, Jodi Forlizzi, and Siddhartha S. Srinivasa. Effects of Robot Motion on Human-Robot Collaboration. *ACM/IEEE International Conference on Human-Robot Interaction*, 2015-March:51–58, 2015.

[13] Minae Kwon, Sandy H. Huang, and Anca D. Dragan. Expressing Robot Incapability. *ACM/IEEE International Conference on Human-Robot Interaction*, pages 87–95, 2018.

[14] Cynthia Breazeal. Social Robots that Interact with People. *An Anthropology of Robots and AI*, pages 72–88, 2008.

[15] Javier Fernandez De Gorostiza Luengo, Fernando Alonso Martin, Alvaro Castro-Gonzalez, and Miguel Angel Salichs. Sound synthesis for communicating nonverbal expressive cues. *IEEE Access*, 5:1941–1957, 2017.

[16] Emma Frid and Roberto Bresin. Perceptual Evaluation of Blended Sonification of Mechanical Robot Sounds Produced by Emotionally Expressive Gestures: Augmenting Consequential Sounds to Improve Non-verbal Robot Communication. *International Journal of Social Robotics*, 14(2):357–372, 2022.

[17] Diana Löffler, Nina Schmidt, and Robert Tscharn. Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound. *ACM/IEEE International Conference on Human-Robot Interaction*, (March):334–343, 2018.

[18] Michael Schmitz, Benedict C.O.F. Fehringer, and Mert Akbal. Expressing emotions with synthetic affect bursts. In *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, pages 91–96. Association for Computing Machinery, Inc, 10 2015.

[19] Helen Hastie, Pasquale Dente, Dennis Küster, and Arvid Kappas. Sound emblems for affective multimodal output of a robotic tutor: A perception study. *ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 256–260, 2016.

[20] Lisa Zahray, Richard Savery, Liana Syrkett, and Gil Weinberg. Robot Gesture Sonification to Enhance Awareness of Robot Status and Enjoyment of Interaction. In *29th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2020*, pages 978–985. IEEE, 8 2020.

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

[22] Takanori Komatsu. LNCS 3784 - Toward Making Humans Empathize with Artificial Agents by Means of Subtle Expressions. Technical report.

[23] Jason Sterkenburg, Myounghoon Jeon, and Christopher Plummer. Auditory emoticons: Iterative design and acoustic characteristics of emotional auditory icons and earcons. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8511 LNCS(PART 2):633–640, 2014.

[24] Richard Savery, Amit Rogel, and Gil Weinberg. Emotion musical prosody for robotic groups and entitativity. *2021 30th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2021*, pages 440–446, 2021.

[25] Richard Savery, Lisa Zahray, and Gil Weinberg. Before, Between, and After: Enriching Robot Communication Surrounding Collaborative Creative Activities. *Frontiers in Robotics and AI*, 8(April):1–11, 2021.

[26] Jacek Grekow. Music emotion maps in the arousal-valence space. *Studies in Computational Intelligence*, 747:95–106, 2018.

[27] James A. Russell. Core Affect and the Psychological Construction of Emotion. *Psychological Review*, 110(1):145–172, 2003.

[28] Richard Savery, Ryan Rose, and Gil Weinberg. Establishing Human-Robot Trust through Music-Driven Robotic Emotion Prosody and Gesture. *2019 28th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN 2019*, pages 4–10, 2019.

[29] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994.

[30] Gentiane Venture and Dana Kulić. Robot Expressive Motions. *ACM Transactions on Human-Robot Interaction*, 8(4):1–17, 2019.

[31] Hiroko Kamide, Yasushi Mae, Koji Kawabe, Satoshi Shigemi, Masato Hirose, and Tatsuo Arai. New measurement of psychological safety for humanoid. *HRI'12 - Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 49–56, 2012.

[32] Nicholas J. Hetherington, Ryan Lee, Marlene Haase, Elizabeth A. Croft, and H. F. MacHiel Van Der Loos. Mobile Robot Yielding Cues for Human-Robot Spatial Interaction. *IEEE International Conference on Intelligent Robots and Systems*, pages 3028–3033, 2021.

[33] Jia Chuan A. Tan, Wesley P. Chan, Nicole L. Robinson, Elizabeth A. Croft, and Dana Kulic. A Proposed Set of Communicative Gestures for Human Robot Interaction and an RGB Image-based Gesture Recognizer Implemented in ROS. 2021.

[34] Peter Grier. Buzz. *Air Force Magazine*, 99(9):87–92, 2016.

[35] Cynthia Breazeal. Emotion and sociable humanoid robots. *International Journal of Human Computer Studies*, 59(1-2):119–155, 2003.

[36] Maitreyee Wairagkar, Maria R. Lima, Daniel Bazo, Richard Craig, Hugo Weissbart, Appolinaire C. Etoundi, Tobias Reichenbach, Prashant Iyengar, Sneh Vaswani, Christopher James, Payam Barnaghi, Chris Melhuish, and Ravi Vaidyanathan. Emotive Response to a Hybrid-Face Robot and Translation to Consumer Social Robots. *IEEE Internet of Things Journal*, 9(5):3174–3188, 2022.