TOWARD GENERATIVE SOUND CUES FOR ROBOTS USING EMOTIVE MUSIFICATION

Ibrahim Syed and Jason Fick

Music Department
Oregon State University
Corvallis, OR, USA
{syedi, fickj}@oregonstate.edu

Brian J. Zhang and Naomi T. Fitter

CoRIS Institute
Oregon State University
Corvallis, OR, USA
{zhangbr, fittern}@oregonstate.edu

ABSTRACT

Sound is an essential yet under-studied way for robots to communicate with humans. When designed well, robot sound can improve aspects of human-robot interaction from social perception to team fluency, but at its worst, robot sound can discourage use of robots altogether. Thus, sound cues must be carefully and intentionally designed. To address this need, we present a system that uses the inherent emotional connotations of musical qualities to dynamically generate emotive sound for robots. Our application utilizes real-time modification of tempo, pitch, scale, and sound brightness to algorithmically generate melodic phrases intended to evoke specific moods or feelings in human listeners. An in-the-wild exploratory study with $N = 26$ participants demonstrated that our generative sounds caused human listeners to perceive the robot as happier and warmer. This effort is a first step toward a planned full system that will democratize the design of music-based emotional communication in human-robot interaction.

1. INTRODUCTION

As robots collaborate and interact with humans, effective methods of robotic expression must be developed and refined for the sake of smooth and conducive human-robot transactions. While robot movement and facial expressions are typically used to enhance human-robot communication, not all robots have the capability to express wide ranges of movement and many lack expressive faces. Therefore, other methods of expression (namely sound) are essential tools for robots to convey information [1]. Our goal is to use sound to begin to explore musification as a way to express emotion in the context of human-robot interaction.

Nonverbal robot sound provides a way for robots to articulate information to humans using audio cues. Transformative sound, nonverbal sound designed specifically to augment robots rather than mechanical noise generated as a consequence of robot operations, can be a way to build trust between robots and humans, and make humans more comfortable during human-robot interaction [2, 3]. Music is a powerful vehicle for conveying mood and emotion [4], and as such, musical sonification is a promising way to develop transformative sound for robots. The term "musification" describes musical sonification: the process of transforming data into musical sound [5]. Musification can be used to sonically express certain emotional states by utilizing the inherent relationship between music and emotion [6]. Emotional musification in particular can be useful for robots that work alongside humans, which may need to communicate intent or specific aural messages such as warning signals [7]. Using the ideas of musical emotion and real-time sonification, we have designed an application that integrates with a robot and automatically generates emotive musical patterns to provide robot actions with emotional connotations. This work examines a robotic system prototype equipped with the ability of emotional musification to convey certain emotions independently as a robot operates.

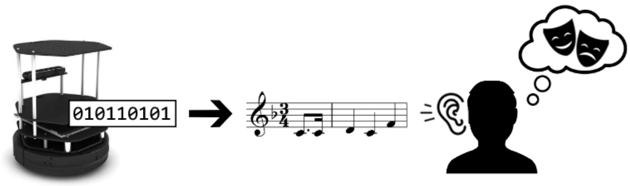

Figure 1: Illustrative image of our system. As a robot operates, the application automatically generates emotive melodies.

Several systems to generate and modify sound cues based on robot data have been developed in past work [8, 9], but exploration of how musical cues can specifically imply emotions and convey information in the context of robot sonification is sparse. Beyond the scope of robotics, though, musical characteristics and their effects on perceived emotion is a well-documented area of study. Research shows that music can be classified by emotion based on qualitative characteristics [10]. These aspects, such as rhythm, spectral dynamics, and harmony can in turn be changed to affect the emotional quality of musical cues [11]. Additionally, work in the field of musification has explored the translation of data into music in depth, but little exists toward the end of applying musification methods to robot sound. We seek to combine these two fields to explore music generation methods for robot sound and how manipulating musical qualities can in turn affect emotional communication in human-robot interaction.

In this paper, we provide an overview of an application we developed to algorithmically generate musical cues for use in robot sound, the idea of which is shown in Fig. 1. We also discuss the methods and techniques used to generate and manipulate these musical patterns, and the design of the study we ran to validate the system. In our study, we had a robot interact with participants in a public space, either generating no sound, happy sound, or sad sound. We recorded the participants' responses and thoughts on their interaction with the robot, specifically gauging reactions to the generated sound. The ultimate goal of our system is to aurally convey discrete emotions using audio cues that are grounded in robot states and actions. Currently, our system implements a process to generate melodic patterns with parameters that change to express distinct *happy* and *sad* emotions. Additionally, it provides a way to switch between these emotive states seamlessly. This early prototype can be connected to robot data streams in future steps for the envisioned end-to-end musification. Our application can benefit roboticists and robot sound designers who seek to improve the sound-based communication (and, more broadly, the overall perception) of their robotic systems.

2. RELATED WORK

Robot sonification and impact of sound in robot expressiveness have been explored in a variety of ways. Mechanical sound generated by a robot during movement may interfere with human interpretation, which justifies transformative sonification as a method of augmenting robot clarity [12]. The actual sounds or instrumentation of sonification is a subject of experimentation as well. Sound sources and their associated timbral content have varying connotations that may imply different ideas or emotions, so sound design is an important variable to consider in expressing robot information. In the context of robot gesture sonification, past research explored audio oscillators, emotive samples, artificial motor sounds, percussive sounds, and musical loops as different ways to sonify robot movement [8]. Another effort focused on sonifying robots using short, non-linguistic audio phrases (such as clicks and whirrs), and found that children interacting with these sounds tended to attribute emotions to them [13]. Song et. al explored this topic using a psychological model to evaluate color, sound, and vibration as emotional parameters in human-robot interaction, and found evidence that varying intonations of generated "beeps" suggested higher or lower levels of emotional arousal [14]. For our system, we chose to restrict our instrumentation to simple waveforms; we are mainly focusing on the parameters of our generated sound rather than the sound itself. Additionally, rather than directly translating data into audio, we opted to explore a method of sound generation that operates independently of a robot's data, but has the capability to react and change dynamically based on incoming data if desired.

Some systems and classification methods exist to map robot movement parameters to audio parameters such as pitch and amplitude (e.g., [9]). While this past study also seeks to improve expressivity in human-robot interaction by way of sonification, it focuses mainly on the relationship between movement and sound qualities. Brock et al. presented a generalized sonification framework which translated movement data from motion sensors into MIDI patterns to transform motion information into sound [15]. Our system does not map movement to sound, but instead isolates musical qualities to explore how generative music can be used as a tool to independently convey emotion in the context of robot sound.

The process of generating melodic content in real time is an integral part of our system and a complex topic of study. While our system uses a random number generator to select pitches that exist within the same musical scale, related efforts have presented alternative methods for generating content using a variety of musical ideas. Some projects have used computing techniques such as machine learning to develop frameworks for automatic music composition [16]. Brown et al. defined several probabilistic algorithms for generative melodic and rhythmic phrases that take into account ideas in music cognition such as note proximity, closure, and goal-seeking [17]. The idea of dynamically altering musical elements to affect the mood of algorithmically-generated melodic content, as we have implemented in our system, has been explored in depth by past work which utilized melodic attraction (i.e., which pitch a listener expects to follow another within a musical phrase) to increase and reduce tension in live-generated music [18]. However, this type of sound generation has not been evaluated in real-world deployments of robots, as we begin to do in the presented evaluation.

Study design varies based on the nature of work in previous implementations. Song et al. studied the expressivity of a robot designed with attributes of a human face by placing it in front of participants and having it repeatedly perform a single expression, recording the emotion that participants thought it expressed [14]. One past study analyzing the speed of a robot's approach and volume of its functional sound had robots travel toward a participant at the end of a hall, and took note of their subjective evaluations of the robot's speed and noise [19]. A study exploring children's responses to non-linguistic utterances for use in robotics involved a humanoid robot emitting speech-like sound, having children listen to its utterances, and evaluating their responses using software that records pleasure, arousal, and dominance [13]. In our study, the robot roams around a university space before approaching a participant, either generating no sound, happy sound, or sad sound.

3. SOUND SYSTEM DESIGN

For our proposed context of sound generation for robots, we needed a sound generation method that had roots in the music technology community and also interfaced with robotic standards. After this selection, we designed and built the system itself, using past related research as a foundation. This section details our decision-making processes related to both of these topics.

3.1. Selection of Pure Data

We designed a generative music model in Pure Data (Pd)[1], an open-source visual programming environment for multimedia. While other platforms such as Max/MSP[2] [8, 9] and SuperCollider[3] [20] have been used for both robot sonification and generative music applications, Pure Data was selected for ease of package distribution and relative simplicity. As opposed to Max/MSP, which requires a paid subscription or license, Pure Data offers many of the same features in a free and open source environment. Pure Data also integrates more smoothly with the platforms we used to transmit data to and from our robot; communication between Pd and Python is easier to set up than communication between Max and Python due to Pd's Fast Universal Digital Interface communication protocol, and Max/MSP lacks Linux compatibility, which is commonly required for robotics software. Furthermore, Pure Data (a visual programming language) is more accessible to sound designers due to its audio signal flow-based structure and does not require as much prior programming experience as a text-centric language such as SuperCollider.

Pd offers a proven framework for live generative music, and its capabilities lend itself well to this system due its strength in dynamically altering sound. Pure Data has seen use in similar sonification frameworks, such as the motion sensor sonification model

[1] https://github.com/pure-data/pure-data
[2] https://cycling74.com/products/max
[3] https://github.com/supercollider/supercollider

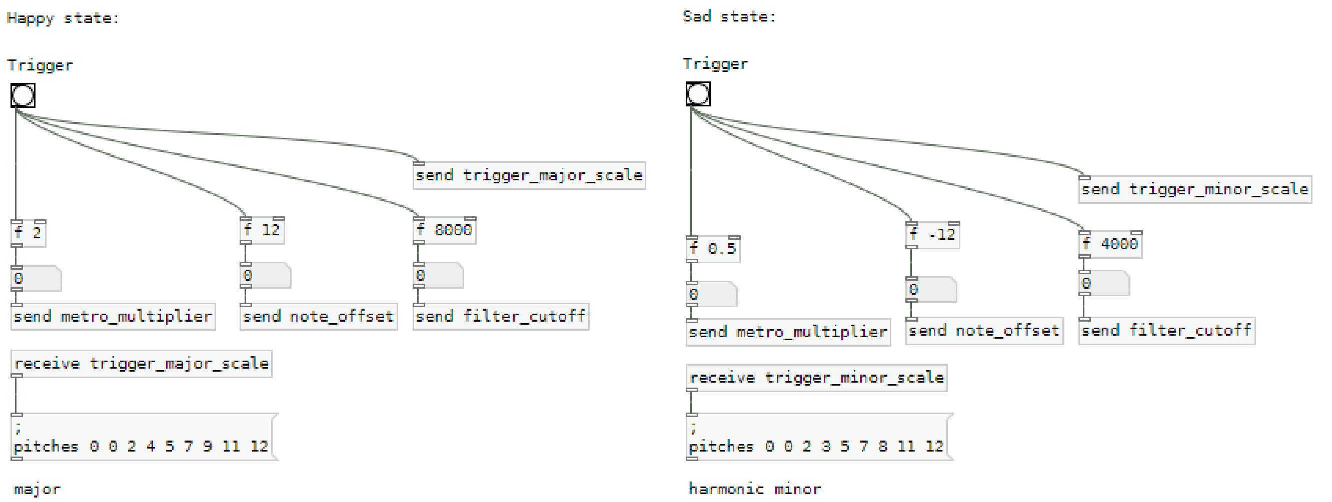

Figure 2: Usage of sends and receives in our patch. Once a state is triggered, its respective values are sent throughout the application to update sound generation.

outlined in [15], which also used the framework for real-time modification of sound properties such as timbre and frequency. Pd provides a way to change the values of numerical parameters across a programmed musical model without interrupting the sound engine with its built-in send and receive messaging system. This messaging paradigm allows programmers to define mutable receiving variables during sound generation (such as pitch or tempo), and then access and modify them during live sound playback by sending values to them from anywhere throughout the program, as demonstrated in Fig. 2. This capability allows our system to not only generate melodies, but also manipulate their sound and expressions seamlessly, preventing hiccups, stuttering, or other unintended sound anomalies that may impede the clarity of a robot's audio. A large benefit of our dynamic system design is that it eliminates the explicit need for roboticists or sound designers to precompose melodic or sampled content to express different robotic states, since all of the necessary sound generation and modification can be done in real-time within our application.

3.2. Sound System Overview

The primary goals of the sonification model are to:

1. Generate a continuous, consonant melodic phrase
2. Parametrize select musical variables to allow for change in emotive quality
3. Create an interface that allows for easy adjustment of these variables without interrupting playback

The model combines a few different components in order to generate musical cues: a basic wavetable synthesizer for sound generation, a randomized sequencer to program rhythmic patterns, and indexed tables of melodic intervals required for major and minor tonalities. The synthesizer has four waveform options (sine, square, sawtooth, triangle), which in turn affect the timbre of the instrument. For the purposes of our case study, waveform type was restricted to sine waves for both happy and sad states, but the capability exists to dynamically change sound wave type to any of the four programmed options. All sound that the synthesizer outputs is routed through a low-pass filter, which removes high-end frequency content past a given cutoff value to make the instrument sound darker or brighter.

Melodic generation is accomplished using a sequencer which iterates through eight triggers in time with a global metronome. This metronome controls tempo, allowing the user to define an eight-note rhythmic pattern at a specific tempo that acts as a starting melodic phrase. At the end of each phrase, each trigger is sent a random signal to either switch state or remain unchanged, which alters the melody by either activating or de-activating certain pitches. When a pitch is triggered, its value is pulled randomly from a table which defines the current scale (either major or minor). The overall structure of our melody generation system is outlined in Fig. 3.

System Melody Generation. Human musicians introduce variations in their performances either intentionally or inadvertently, which can greatly affect the emotional qualities of a piece, but computer-generated music has the potential to sound dull and stale without variation [21]. With the goal of conducive human-robot interaction in mind (we want human listeners to maintain awareness of the robot's sound cues, not tune them out), the generative system needed a way to address this concern. Our patch introduces rhythmic variation by using a random element that slightly alters the note pattern at the end of each musical bar. Additionally, restricting the magnitude to which rhythm is altered (by reducing the probability that each pitch is randomly triggered on/off at the end of a bar) ensures that newly generated note patterns retain some of the rhythmic qualities of their preceding pattern, avoiding abrupt or jarring changes. The nature of our algorithmic melodic generation lends itself to variation as well; every time the synthesizer is triggered, it plays a random note from whichever scale is active. When uninterrupted, the patch's generated melody will remain in one tonality, but the pattern in which the scale's notes are selected is constantly in flux. This allows for melodic diversity while retaining consonance by restricting note selection to a single tonality.

Parameterization in the System. There is strong precedent for the idea that music can convey specific emotions with intentionality, and that the values of musical parameters such as pitch and tempo

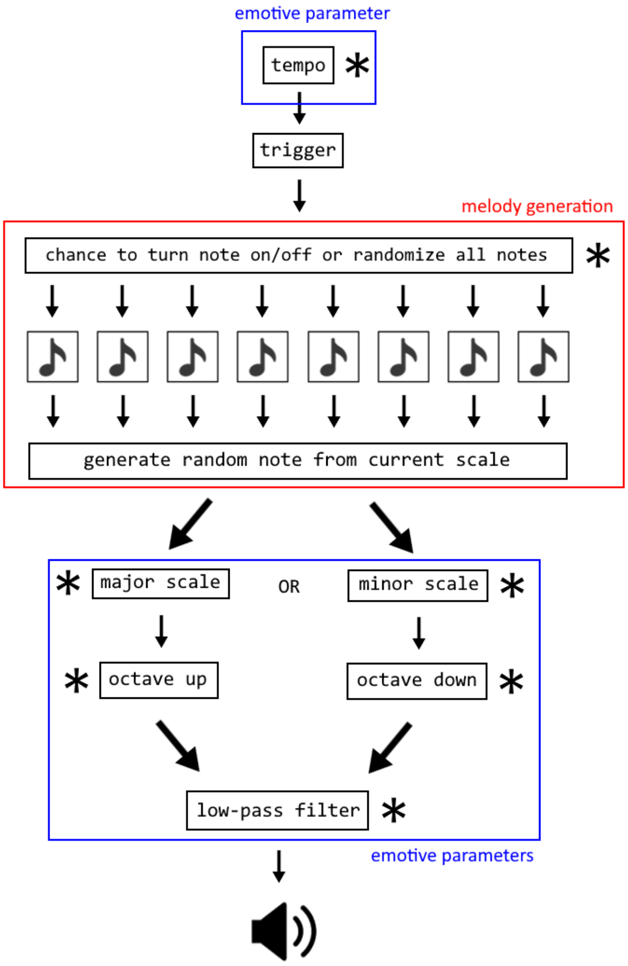

Figure 3: A simplified overview of our melodic generation method. The system is comprised of two main parts: melodic generation and emotive parameter manipulation. All variables marked with an asterisk (*) are values that change based on the selected emotional state.

can suggest certain emotions or convey certain signals [7, 22, 23]. Rhythm, timbre, harmony, and spectral features all play a role in expressing musical emotion [10]. Accordingly, one primary function of our application is to allow for manipulation of these musical parameters, either manually by the user or automatically based on input data, to change the emotive quality of generated sound in real time.

The main parameters being modified in our system are scale, tempo, octave, and cutoff frequency of a low-pass filter. Each of these individual variables can be manipulated in such a way that increasing or decreasing them changes some emotive quality of the melody being generated, and changing all of these variables at once can entirely change its emotional connotation [24]. For example, modulating the scale of a melodic phrase can represent a change in the perceived emotion of the material. This is reflected in the generative patch, which produces melodic material in either major or harmonic minor keys. Traditionally, music in major keys is associated with a happier mood, and vice versa for minor tonalities (although these are not the sole distinguishers of musical emotion). A similar dichotomy is found with tempo; slower tempi are associated with sadder connotations, while faster tempi seem happier. As such, we added functionality to vary tempo in our application. Additionally, the presence and loudness of higher frequencies in a sound is often associated with brightness, while lower frequency content can be perceived as darker [11]. Considering this, we implemented octave offset and low-pass filter controls to modify the brightness (or darkness) of the produced sound; lower octaves contain lower frequencies than higher octaves, and the low-pass filter allows for the direct removal of high frequency content in the sound. It should be noted that the correlations between these sound parameters and happy or sad emotions are not strict rules, but provide a necessary basis and simple starting point for our generative sound framework.

Investigated Sound States. Before any parameter modulation is applied, the Pure Data patch generates a default melodic state. Default values include a starting pitch (C4), tempo (100 BPM), tonality (major), and low-pass filter cutoff (6000 Hz). From this starting neutral state (where each starting parameter can be adjusted to liking by the user), the patch interfaces with these variables by applying pre-defined transformations to each one based on a selected emotion, happy or sad. These two emotive states change the value of each parameter in a way guided by the emotional connotations discussed above. Specifically, the happy state parameters include an octave offset of +1 (one octave above the starting pitch), a major tonality, double tempo (the melody plays at double the default speed), and a low-pass filter with an 8000 Hz cutoff. The sad state includes an octave offset of -1, harmonic content in a minor (harmonic) tonality, half tempo, and a low-pass filter with a 4000 Hz cutoff. These states and the parameters associated with them are represented in Table 1.

Currently, the emotional state of generated audio cues is selected manually by the application user, and in turn emotive parameters in the application are changed to reflect the values outlined above. However, within our framework, the capability exists to attach triggers to both emotional states that would allow for state to change automatically based on programmer-defined robot behavior (such as in [25]). For example, a robot's sound could switch to happy when it senses a person nearby, or sad if it becomes trapped in a corner.

Happy and sad emotive states were chosen specifically due to their ease of recognition and pre-existing correlations to musical parameters that are easy to modify, such as scale [11]. Additionally, emotional models such as Russell's circumplex model place discrete emotions of sadness and happiness at opposite ends of one key emotional axis [26]. Therefore, happy and sad musical cues should be easy to distinguish from each other and make for a good proof of concept choice for our system.

	happy	sad
scale	major	minor (harmonic)
tempo multiplier	$\times 2$	$\times 0.5$
octave offset	$+1$	-1
filter cutoff	8000 Hz	4000 Hz

Table 1: The values of audio parameters (rows) for happy and sad emotive states (columns).

4. SYSTEM DEMONSTRATIONS & RECEPTION

We conducted an in-the-wild exploratory study to understand how people perceived the happy and sad emotive states, in addition to a control state without any transformative sound. Our university ethics board approved the study procedures.

4.1. Methods

Robot hardware. We implemented the model on a TurtleBot 2 robot [27]. This robot consists of a non-holonomic mobile base, atop which modular layers of shelves and hardware that can be added. The TurtleBot is a good choice for exploratory robot sound research because of its flexibility and reasonable representation of modern robotic systems. For example, it is easy to integrate custom speakers with this platform; in prior work, we designed a custom speaker system with external power sourcing and a 3.5mm audio input, and for the present work, we used a commercially-available USB speaker system as part of the custom hardware stack. This research hardware also included a Raspberry Pi for processing and 2D lidar sensor for supporting navigation. This mobile robot also parallels commercial robotic systems used in the service industry, especially in delivery applications. In our exploratory study, we outfitted the robot with a candy bowl to give a similar context to interactions with the robot.

Procedure. We deployed the robot in an approximately 3×3m square, bounded by tape, inside the lobby of a university residence hall. The study took place over three days from 10AM to 3PM on each day. The robot displayed one sound profile on each day, as further described below. Signs in English and Spanish described the study and the informed consent process to participants, who could enter the space to consent and participate. Figure 4 shows the study space and robot.

During the study periods, the robot's ambient behavior was proceeding back and forth in a straight line, turning to each side at the endpoints. When participants entered the study space, the robot detected the participant using its 2D lidar sensor. The robot then turned to face and approach the participant, stopping near the participant's feet. A sign on the robot asked "Would you like a piece of candy?" Participants could take candy freely from the bowl on top of the robot.

After each participant interaction with the robot, a study facilitator asked the participant if they would like to complete a brief recorded interview. Consenting participants then completed this interview and were thanked for their time and participation.

Independent variable. The central manipulation in our study was the robot sound profile. Using the sound system proposed in Section 3, as well as a control state, we established the following three conditions:

- *Control:* no sound is added to the robotic system. Visitors only heard the natural sounds emergent from the robot hardware (e.g., sounds from motors and gears).
- *Happy:* the robot displayed the happy emotive state from our proposed sound system. The natural sound profile of the robot was also audible.
- *Sad:* the robot displayed the sad state from our proposed sound system. The natural sound profile of the robot was also audible.

Figure 4: The robot (left) and a facilitator in the study space.

The conditions were presented on consecutive days with the order in the list above. A reminder of the sound system settings for the happy and sad states appears in Table 1, and recordings of each robot sound profile in the study environment can be found in the attached media [28].

Measurement. The main data collected from participants was interview information, which was conducted in a semi-structured format. This core questions from the interview are listed below:

- What made you want to interact with the robot?
- Did anything about the robot stand out to you, such as how it looked, moved, or sounded?
- How did you feel about how the robot sounded? Did it seem happy, sad, or neutral?

The question involving our study conditions was deliberately placed third, so as not to influence initial feedback on the system. This final query served in part as a manipulation check, in addition to supplying more information about the experience of interacting with the robot.

Participants. Over the course of the study, approximately 50 participants interacted with the robot. Among these participants, 31 individuals completed an interview about their interaction with the robot. Because of the varying flow of people through the study space, interviews included between one and three participants at one time. No demographic information was collected on participants, but based on our informal observations, most of the visitors appeared to be young adult university students.

4.2. Results

Our analysis of this exploratory study centered on a sentiment analysis on the collected interview data. A trained coder looked for positive, neutral, and negative comments related to the robot sound specifically in the interview data.

Nine participants completed interviews during the control condition day. After interacting with the robot, participants in none of the eight interviews brought up sound of their own volition. When asked specifically about the sound, participants responded neutrally to negatively. One participant "...did not notice that it had a sound," while participants that did notice described the sound as between "...fine, there was nothing really annoying about it" and "definitely on the noisy-ish end." No participants ascribed a positive or negative emotion to the robot's sound profile.

Thirteen participants spoke with us during the happy condition day of the deployment. Adding sound meant to convey a happy emotive state led to differing impressions of the robot and its sound. Without prompting, participants in four of nine interviews mentioned the sound before the third core question. The sound generally succeeded at conveying a happy state, with responses including statements such as "the beeping sounds very happy," "...the dinging, I liked that; it was pleasant," and "[it sounded] happy, absolutely." While six of the thirteen interviewees perceived the happy state, one participant noted that "it sounded good, but noisy," while another person found that "it just sounded neutral and robot-like." Thus, the majority of participants interpreted the sound as intended.

Nine participants completed an interview on the sad condition day. Largely, rather than being perceived as intended, the sound meant to convey a sad emotive state led to perceptions of the robot as slightly positive of neutral. Participants in three of nine interviews noted the sound before being asked the core interview question on sound, and participant impressions ranged from neutral to positive. Most participants (four of the nine interviewees) responded with a variation of "neutral, I think...it didn't sound happy or sad." Two participants contributed responses such as "happy; its a nice little song" and "kind of calming." One participant found the sound confusing, stating that "the beeps don't really make sense too much. Is it saying 'hello,' is it saying 'do you want something?"' The sad sound condition did seem to convey a different meaning than the other conditions, but it did not convey sadness to participants.

5. DISCUSSION

In the early evaluation of our sound generation system, the *happy* sound profile condition was successful, with most participants perceiving it as planned. It also tended to stand out to participants with clear intent (i.e., as a sonic profile meant to convey positive affective state). The *sad* sound profile condition was interpreted in a broader range of ways. Generally, it was seen as happy or neutral. The specific sad sound setting used in this study was less successful than the happy setting on two fronts: it seemed to be less noticeable and incorrectly interpreted. One participant even expressed explicit confusion as to the meaning of the sound, puzzling over whether the sound was a salutation, prompt, or something else. Both of these conditions were considered relative to a *control* condition with no added sound profile. Interestingly, both the happy and sad conditions tended to be perceived as happier than the control condition. This may mean that a robot with a broad range of sonic profiles will be perceived more positively than a robotic system with no added sound.

The related literature on robot affect shows some precedents for the observed positive skew in perception of robot emotional display. For example, prior work by a subset of the author team showed that people generally perceive emotional robot behaviors more positively than expected [29]. In other work, recognition rates for negative or neutral valence and low energy sounds (e.g., sound to convey hesitation or questioning) were likewise lower than recognition rates for other messages [30]. As in previous work, people also related sonic profiles to sounds from their own personal lives and experiences. For example, participants brought up parallels to music played at Halloween or in Kahoot, an interactive learning game.

Strengths of this work include the system's ability to dynamically change the emotive qualities of a synthesized melody given specific musical parameters (i.e., octave, scale, tempo, and filtering). Our generative sound system is easier to modify, more responsive, and more dynamic than audio files, enabling easy emotional sound synthesis for robot sonification. Further, we successfully deployed the sound system on a real robot. Compared to existing frameworks for robot sound, which typically involve pre-recorded sound that is triggered by a small number of specific events, the proposed sound generation system gives us a broad variety of control over sound, sound triggering, and sound adaptation. For example, the system affords dynamic changes, control over a variety of parameters, and the mapping of sound triggering to existing robotic sensor states. In coming steps, we plan to open-source this work to help others develop sound for robots smoothly using our platform. Broadly, we aim for our system to give roboticists and sound designers a centralized and convenient way to generate and manipulate melodies to change their emotional connotations and thus influence interactions between robots and humans.

Limitations of the current system include the fact that it can only generate happy and sad states at present. A potential expansion for the system could involve introducing new emotive states. To this end, Russell's circumplex model of emotion provides a guideline for potential emotional states to be sonified [26]. The current generated sounds are also relatively simple in terms of rhythm, tempo, and melody. Additionally, our current implementation of melodic generation chooses notes randomly from a scale; traditionally, in western music, some notes in a scale are more significant than others, and effective melodies are structured in a way that takes into account the intervals and significance between notes. This could be a contributing factor towards participants' difficulty in linking our musified sound with the intended perception of sadness. Implementations of more complex rhythmic and melodic generative algorithms (such as the examples in [17]) can increase the intricacy of our musical cues, and aid in conveying the intended emotions of our melodies. Exploring the use of chords in sound generation can provide further clarity in emotional expression as well, as opposed to just melodic phrases. We believe that even with minimal melodies, this chordal aspect might assist in the successful conveyance of sad affect. On the evaluation side, ambient noise in the study space may have influenced the audibility of the robot's sonic profile. The robot's sound intensity and pitch could be adjusted in the future to account for background noise. We also studied a single order of conditions and used only interview data to evaluate the system. A longer study with counterbalancing and additional measures would help us to gather a more reliable and complete picture of the system.

In our *future work*, we will update the settings for recommended default happy and sad emotional states based on the evaluation results, in addition to designing further emotional profiles for the sound system. We also plan to develop a framework for activating emotional state triggers based on robot sensor data and perform a broader range of evaluations of the sound generation system. We have explored this previously in [25], where robot sensor data was used to dynamically trigger different audio cues; a system similar to this can be implemented in our program, allowing for emotional state to change in response to a continuous data stream. In a follow-up study, a broader range of participant data could be collected, including demographics and backgrounds in musicianship which could affect participants' perceptions of our sounds. The proposed type of sound generation can augment communication in robotic systems, including robots with varying levels of nonverbal expression capability via other communication modes. Clear sound cues for robots are essential for fluent robot communication, and our exploration of the intersection of music cognition and robot sound can help to push robot sonification to a new level.

6. ACKNOWLEDGMENT

We thank Melissa Greff for accommodating our study in the International Living and Learning Center.

7. REFERENCES

[1] C. Bethel and R. Murphy, "Survey of non-facial/non-verbal affective expressions for appearance-constrained robots," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, pp. 83 – 92, 2008.

[2] D. Moore and W. Ju, "Sound as implicit influence on human-robot interactions," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2018, pp. 311–312.

[3] B. J. Zhang, N. Stargu, S. Brimhall, L. Chan, J. Fick, and N. T. Fitter, "Bringing wall-e out of the silver screen: Understanding how transformative robot sound affects human perception," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 3801–3807.

[4] P. Juslin, "From everyday emotions to aesthetic emotions: Toward a unified theory of musical emotions," *Physics of Life Reviews*, vol. 10, 2013.

[5] A. D. Coop, "Sonification, musification, and synthesis of absolute program music," in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2016.

[6] A. Godbout, I. A. T. Popa, and J. E. Boyd, "Emotional musification," in *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion (AM'18)*. New York, NY, USA: Association for Computing Machinery, 2018.

[7] J. Fagerlönn, "Expressive musical warning signals," in *Proceedings of the International Conference on Auditory Display (ICAD), 2007*, 2007.

[8] L. Zahray, R. Savery, L. Syrkett, and G. Weinberg, "Robot gesture sonification to enhance awareness of robot status and enjoyment of interaction," in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2020, pp. 978–985.

[9] J. Bellona, L. Bai, L. Dahl, and A. LaViers, "Empirically informed sound synthesis application for enhancing the perception of expressive robotic movement," in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2017, pp. 73–80.

[10] Y. Song, S. Dixon, and M. Pearce, "Evaluation of musical features for emotion classification," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2012.

[11] P. Juslin and P. Laukka, "Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening," *Journal of New Music Research*, vol. 33, pp. 217–238, 2004.

[12] E. Frid, R. Bresin, and S. Alexanderson, "Perception of mechanical sounds inherent to expressive gestures of a nao robot - implications for movement sonification of humanoids," in *Proceedings of the Sound and Music Computing Conference 2018 (SMC2018)*, 2018.

[13] R. Read and T. Belpaeme, "How to use non-linguistic utterances to convey emotion in child-robot interaction," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2012.

[14] S. Song and S. Yamada, "Expressing emotions through color, sound, and vibration with an appearance-constrained social robot," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2017.

[15] H. Brock, G. Schmitz, J. Baumann, and A. O. Effenberg, "If motion sounds: Movement sonification based on inertial sensor data," *Procedia Engineering*, vol. 34, pp. 556–561, 2012.

[16] Weber and S. Alegre, Torresen, "Parameterized melody generation with autoencoders and temporally-consistent noise," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2019.

[17] A. Brown, T. Gifford, and R. Davidson, "Techniques for generative melodies inspired by music cognition," *Computer Music Journal*, vol. 39, pp. 11–26, 2015.

[18] R. Nikolaidis, B. Walker, and G. Weinberg, "Generative musical tension modeling and its application to dynamic sonification," *Computer Music Journal*, vol. 36, pp. 55–64, 2012.

[19] M. Lohse, N. van Berkel, E. M. A. G. van Dijk, M. P. Joosse, D. E. Karreman, and V. Evers, "The influence of approach speed and functional noise on users' perception of a robot," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 1670–1675.

[20] M. S. Schwenk and K. O. Arras, "R2-d2 reloaded: A flexible sound synthesis system for sonic human-robot interaction design," *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 161–167, 2014.

[21] R. Bresin and A. Friberg, "Expressive musical icons," in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2001.

[22] E. Schubert, S. Ferguson, N. Farrar, and G. E. McPherson, "Sonification of emotion i: Film music," in *Proceedings of the International Conference on Auditory Display (ICAD)*, 2011.

[23] A. Gabrielsson, "Strong experiences with music: Music is much more than just music," *Strong Experiences with Music: Music is much more than just music*, pp. 1–512, 2012.

[24] E. Schubert, "Modeling perceived emotion with continuous musical features," *Music Perception*, vol. 21, pp. 561–, 2004.

[25] B. J. Zhang, N. Sigafoos, R. Moffit, L. Adams, I. Syed, J. Fick, and N. T. Fitter, "Sonifyit: Towards transformative sound for all robots," accepted for publication in the IEEE Robotics and Automation Letters (RA-L), 2022.

[26] J. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980.

[27] Open Source Robotics Foundation, Inc., "TurtleBot2." [Online]. Available: https://www.turtlebot.com/turtlebot2/

[28] Anonymous, "Media for "Towards Generative Music Cues for Emotive and Responsive Robot Sonification"," 2022. [Online]. Available: https://www.youtube.com/watch?v=W0g_9gKk5x8

[29] L. Chan, B. J. Zhang, and N. T. Fitter, "Designing and Validating Expressive Cozmo Behaviors for Accurately Conveying Emotions," in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Aug. 2021, pp. 1037–1044.

[30] J. Fernandez De Gorostiza luengo, F. A. Martin, A. Castro-Gonzalez, and M. A. Salichs, "Sound Synthesis for Communicating Nonverbal Expressive Cues," *IEEE Access*, vol. 5, pp. 1941–1957, 2017.