A Case Study on Robot Sound Design for a Sidewalk Delivery Robot

Brian Zhang[1], Ibrahim Syed[1], Jason Fick[1], and Naomi T. Fitter[1]

[1]Oregon State University (OSU), Corvallis, OR, USA
{zhangbr, fittern}@oregonstate.edu

Abstract. Robot sound influences aspects of human-robot interaction (HRI) from the way robots are perceived socially to the adoptability and even monetary value of these systems. Accordingly, our past work (and the work of others in the HRI community) has sought to empirically investigate robot sound and propose design tactics for successfully designing and incorporating new sound use in robotic systems. This paper presents a case study of using these collective past works to design and evaluate robot sound as used on a real delivery robot (Daxbot's Dax robot) in its day-to-day commercial operations. We present the steps of our design thinking process for incorporating new and beneficial sound into the Dax robot, followed by evaluations of these designs in the wild and through a broad online survey. In-the-wild users ($n = 6$) found the newly designed sounds to be agreeable, and more participants preferred the new sounds compared to past built-in sounds of the robot. The online survey-based study ($n = 75$) supported these initial findings and showed that the added sounds also yielded higher perceived robot competence and purchasing interest compared to a base no-sound condition. Overall, this work offers evidence that our team's past robot sound work (often conducted in controlled environments with research robots) has the potential to transfer to real-world use settings and commercial robots.

1 Introduction

The way a robot sounds matters, from sound that naturally emanates from low-level mechanisms of a robot to expressive sound added atop a robot's natural sonic profile via speakers. For example, quieting the sound of robot mechanisms can result in a robot seeming more competent and less discomforting [23]. Adding character-like sound to a robot's behaviors makes the agent seem more socially warm and more competent [26]. We can even measure differences due to sound at the value level, where added character-like sound results in a significantly higher suggested price for sonically augmented robotic systems [24]. The existing (and growing) corpus of robot sound work paints an impressive picture for the importance of this topic, but most work in this area to date (with a few notable exceptions, e.g., [11,12]) draws conclusions based on online video-based studies. Based on this tendency, we became curious about the value of other types of work in the robot sound space, and documented a recommended design methodology for successful robot sound [22]. The work presented here follows the previously established design process to establish and evaluate new sounds for a commercial service robot in a mix of in-the-wild and online evaluations. We propose that this case study is a meaningful foundation for translation science in the domain of robot sound.

In this paper, the described design process and evaluations center on a collaboration with Daxbot, a robotics company currently engaged in food delivery in Oregon. The author team includes roboticists and music technologists with sound design experience who worked together along with Daxbot to use a design thinking process to update the sound profile of the commercial Dax robot and conduct two major evaluations rounds to help validate the resulting sounds. Our main research goals were to *design appropriate sounds for a commercial service robot* and *demonstrate the concrete impact of the sounds via feasible evaluations.* We describe the Dax robot and other key background (such as grounding on the design thinking process) in Section 2. Section 3 describes the design thinking process for updating the sound profile of the Dax robot, and Sections 4 and 5 detail evaluation efforts for the newly designed robot sounds. The discussion and conclusions of the work are in Sections 6 and 7. Overall, this work's contributions include: 1) the modeling of an exemplar robot sound design process in action (as proposed in theory in a past paper [22]), and 2) the proof-of-concept evaluation of the resulting robot sounds.

2 Background

This section covers key related information about robot sound, the Dax robot, and the design thinking process.

Robot sound is a relatively new but growing research area that covers ideas from verbal speech to non-linguistic (or even unintentional) use of sound by robots. In the case of our focus, we considered speech, but ended up focusing mainly on the design space of nonverbal robot sound such as consequential sound, which naturally emerges from a robot's mechanical makeup and interactions with the environment [7,9], and functional, transformative, and emotional sound (i.e., sound meant to provide information, change a robot's natural sound, and convey feelings using approaches from gibberish speech to music, respectively), which can be added to a robot's natural sound profile via speakers to alter the sound in some way [21]. We only had the ability to alter the sound played by the robot (not the mechanical makeup) during our collaborative design process, so in this paper, we focus on the latter sound types. There is also generally more work focused on these added sound types (e.g., [3,18,14,15,12]). Typically, these works demonstrate benefits of augmented sound profiles generally or benefits of specific types of sound, as in the cases of the past work mentioned at the start of Section 1. The results presented in the current paper augment the body of work on potential impacts of added robot sound with increasing ecological validity.

The *Dax robot* is a outdoor mobile robot with interactive features such as a two-degree-of-freedom neck, head, eye display, and speakers, as shown in Fig. 1. Based in Philomath, Oregon, Dax operates in Philomath and Monmouth to perform food and grocery delivery [6]. At the time of our interactions with Daxbot (periodically during 2022-23), we learned that during the delivery process, Dax alternated between autonomous control and operator control modes.

Fig. 1. A Dax robot driving on the sidewalk facing the camera

In *autonomous control mode*, Dax could complete all tasks including pick-up, driving, and drop-off; during the process, Dax performed selected nonverbal communications such as nodding and blinking. During autonomous control mode, Dax did not employ any sounds. In *operator control mode*, operators had the option to trigger four different nonverbal sounds: a positive sound, a neutral sound, and two negative sound options. Noting that these sound options were limited, we sought to understand new opportunities for Dax sound design and to deploy and evaluate our proposed sounds on an actual Dax robot.

We used the *design thinking process* in this paper, as encouraged by our own past framework in [22]. The design thinking approach involves five main steps: empathize, define, ideate, prototype, and test [5]. This approach is consistent with other past human-robot interaction (HRI) work, especially efforts that seek to design robots effectively for real use cases and/or use by special populations. For example, other research teams used design thinking in the development of the Vizzy robot, which supports exercise via augmented reality games [10,17], and the Stevie robot, a socially assistive robot for retirement communities [8]. In the case of this paper, using an established design process for introducing robot sound into real-world use cases is still a relatively novel idea (used by few past works, such as [12]). We believe that the clear documentation of our process, including early foundational steps to support the sound design, can help other HRI researchers to boost their own design and real-world evaluations of nonverbal robot expression.

3 Design Thinking-Based Robot Sound Design

In our own past work, we engaged with roboticists and sound designers to learn more about robot sound design needs and processes, and we concluded by proposing design thinking process steps for successful robot sound design [22]. Accordingly, at the start of the work with Daxbot, we followed our own established best practice tips, aiming to understand different perspectives important to the robot sound experience, ideate possible sound designs in a cross-disciplinary way, work with the company to deploy needed system elements, and assess the system prototypes through rounds of evaluation. The following subsections detail the design thinking process for the Dax robot sound design, as well as details on the final produced sounds.

3.1 Design Thinking Process

Empathize: As part of the empathizing process (specifically, to take the perspective of incidental passers-by and direct end users who order food from the robots), research team members (including one roboticist and one music technologist with sound design experience) made five trips to Philomath to order food from Dax. During each trip, we walked to the origin point of the delivery to observe the robot throughout the entirety of the process (i.e., gaining perspective on the pedestrian passer-by experience, not just the experience of those who place an order with the robot). Initial trips yielded different experiences than later trips; the first trip, for instance, led to an in-depth expressive interaction with head motions and facial expressions (such as "heart eyes"). However, later trips led to brief interactions with a relatively minimal nod from the robot upon confirmation of delivery and receipt. Through interactions with Daxbot, we learned that (at the time) this difference in user experience was due to whether Dax was operating autonomously being controlled by a human operator. From the customer's point of view, however, Dax appeared to be particularly social or asocial, with asocial behavior potentially leading to a disappointing interaction. Even during the in-depth interaction, we noted that the robot produced many motions but no intentional sounds, instead producing only consequential motor and mechanical sounds as the robot's neck moved.

Define: Based on the mock user interaction experiences, we defined the most promising interaction of focus as the customer drop-off interaction. We considered that there are additional users beyond those considered in our perspective-taking exercise, such as food service worker interactions loading food into the robots; however, as food service workers interact repeatedly during a single day and could potentially grow fatigued by the sounds, we opted out of using sound to augment this interaction. We also steered away from seeking to augment interactions with passers-by, since these encounters are very brief and hard to rapidly evaluation in in-the-wild settings. For the identified delivery interaction in particular, we defined the points in the customer interaction during which the robot would begin to act or transition to the next phase as key landmarks. For these time points/windows, we considered both functional sounds that could inform the customer that the robot had acknowledged some information (e.g., a QR code for delivery confirmation or a thumbs-up for receipt confirmation) and transformative and/or emotional sounds to potentially mask motor sounds that often accompany robot motion.

Ideate: In our broad ideation about relevant robot sound design, we considered forms of sounds including: transformative sounds that would cover up consequential sounds during driving, neck motion, and compartment opening/closing; speech, vocables (nonverbal vocalizations, such as humming or exclamations such as "ah!" [1]), and vocable-like musical sounds; and music. At the time of our investigation, Dax did not have the inherent ability to understand a customer's speech in autonomous mode, so we ruled out verbal speech to avoid introducing inflated expectations of the robot. Furthermore, since Dax's head and neck motions were relatively fast and exaggerated, we decided against subtle transformative sound. The small library of existing Dax sounds included mostly vocables. Due to the nature of these existing sounds, combined with the inherent similarity of the Dax robot to cinematographic robots like WALL-E, we ruled out unmodified vocables as a focus and instead sought to expand the robot's nonverbal expression capabilities with vocable-like musical sounds and musical sounds specifically.

Prototype: Using Ableton Live, a digital audio workstation [19], the music technologists on our team began prototyping sounds with the stock instruments and provided synthesizers. First, we laid out simple melodic phrases: two or three short notes in rapid succession, meant to imitate the sounds of robots in media such as WALL-E and R2-D2. Then, we designed three separate synthesizer patches using Ableton's stock Operator and Wavetable plugins. The first (Operator plugin) was an FM patch meant to somewhat replicate the 'gritty'/'metallic' feel of Dax's original sound set. The second, in Wavetable, utilized one of the stock 'formant' Wavetable presets to roughly replicate the sound of human vocal phrases. The last was another Wavetable plugin that layered two different sine waves. After critique and deliberation, we ultimately decided to continue prototyping with the formant Wavetable patch. The formant patch offers voice-like qualities without a high risk of being conflated with an actual vocal sample, and it is musical enough to avoid the potential problems associated with utterances and voice synthesis (e.g., inflated expectations of robot capabilities).

With Dax's base voice clarified, we created phrases for the robot to say. To accomplish this, Ibrahim (one of the team members) first recorded himself saying a set of phrases, such as "duuuude...", "ta-da!", and "yay!" Then, using the monophonic Wavetable patch, we replicated the pitch shifts from the recorded phases on the piano roll. We also modified several parameters to bring more vocal quality to the samples: cutoff filter frequency, detune, and amplitude envelope release. Drawing on our expertise from past experience in robot sound design, we roughly categorized the samples into positive- and negative-valence groupings. For example, "duuuude..." has negative valence and is meant to express disappointment, whereas "yay!" has a clear positive valence. We used filter frequency and detune to make negative valence sounds darker and more dissonant; positive valence sounds were made brighter and did not contain detune. Additionally, for many of the positive-valence sounds, we created melodies using mainly octave and major third intervals to make them as musically consonant as possible (and thus more pleasant for the listener) [16].

For ease of integration for this first sound integration with the Dax robots, we opted to use the existing sound playback system of the robot to avoid the need for complex code additions to the robot's proprietary codebase. This playback system plays .wav files on command. (For more adaptive future sound variations, readers can consider using Pure Data within our SonifyIt tool [25] to play back sounds with live variations, to minimize interaction repetitiveness.)

Test: In early testing of the produced robot sounds, we conducted multiple rounds of internal tests through visits to Dax. First, this testing focused on iterate on sounds as they sounded when actually played back through Dax's speakers. Notably, several sounds required changes to accomplish the desired communication effect when played outdoors on Dax, compared to when played on headphones. When played on Dax, due to a loss of lower frequencies, the sounds at first felt emptier and more structurally weak, especially when compared to the original sound set developed by Daxbot. In addition, we observed that some of the melodies intended to be positive ended up falling flat or not conveying discernible emotion when presented on the robot itself. To address these observed challenges, most sound phrases had their length slightly extended, a sub bass was added to the synthesizer, and the melodic intervals on some of the phrases were modified. After this adjustment, we returned to Daxbot headquarters and successfully conducted an experimental deployment of the new sound set during a robot delivery in the wild. Following this test, small adjustments were made to the timing and placement of sound cues in the codebase. Major evaluations of the sounds are covered in dedicated sections (Sections 4 and 5).

3.2 Sound Design Results

The final sound-based delivery interaction, which includes four custom robot sounds, progressed as follows:

1. Dax arrives at the customer's specified located. Here, "arrival.wav" is played.
2. The customer shows Dax a QR code provided by the ordering process.
3. If the QR code is valid, Dax nods. Here, "yes.wav" is played.
4. Dax opens its compartment. Here, "ta-da.wav" is played.
5. The customer retrieves their items.
6. The customer gives a thumbs-up to Dax.
7. Dax nods. Here, "yay.wav" is played.
8. Dax closes its compartment.

Eight additional sounds were created for possible branches in the progression above, such as the scanning of an incorrect QR code or the robot being impeded by a pedestrian, for a total of twelve novel sounds; one of these sounds ("downward-no") was later used in the online survey-based study. Sounds primarily followed transition states, confirming to the customer that Dax has seen something or is doing something. In addition, sounds generally coincide with (and partially mask the sound of) operations that produce consequential sound. All sound samples are available in [20].

Overall, we expanded the library of communicative sounds available to the Dax robot from four to 12 sound samples, which can allow for some variation even without next-level techniques (e.g., Pure Data and live synthesis). These end products could be used during autonomous operation and/or triggered by human operators during the delivery process, as further evaluated below.

4 Initial Evaluation: In-The-Wild Dax Assessment

As a first more holistic assessment of the new sounds, we conducted an in-the-wild evaluation in Monmouth, Oregon to examine end users' responses to the Dax delivery interactions when augmented with the new sounds during real orders. The following subsections detail the methods for collecting data during these evaluations and the results of the evaluations, which center on anecdotes from interviews. This process was approved by the Oregon State University Institutional Review Board under protocols #HE-2023-186 and #IRB-2019-0481.

4.1 Methods

We sought to collect data from real deliveries with the new sounds while maintaining minimal invasiveness to the delivery process and overall user experience. Accordingly, our approach to collecting data during these interactions involved three methods:

1. After an order, customers received a link to a survey that was designed by our research team. Anyone who completed this survey received 5 USD in compensation.
2. During orders without added sounds, we observed customers from the Dax operator station to form a baseline understanding of customer behavior.
3. During orders with added sounds, we accompanied the robot and conducted a brief semi-structured interview after order completion.

Participation in the survey was extremely low (with only one response) and observations of the order deliveries yielded little rich behavioral information (as most end users were simply focused on taking their order and continuing with their day). Notably, the single survey response showed appreciation in response to the question: "What part(s) of your interaction with Dax stood out to you most or most strongly influenced your responses throughout the survey?": "...His little noises sound like words I understand. He brings me joy!"

Accordingly, the following results focus mainly on the slightly richer input from the semi-structured interviews. We conducted six of these in-person post-delivery interviews, as further discussed below.

4.2 Results

Anecdotally, observations of the deliveries revealed that with the new sounds, interactions remained fairly short (like with the past interaction design), although participants occasionally expressed happiness and verbally thanked Dax during deliveries. The aim of gaining direct information about the existing day-to-day delivery interactions proved difficult without experimenter intervention, with more direct data elicitation protocols showing better results. In direct in-person information elicitation, however, limits on the study facilitator's time and physical location (e.g., the need to be at each delivery interaction) prevented larger-scale data collection.

Interviews showed that all participants liked the robot's sounds in general. For example, users commented that "he makes kind of like little WALL-E-like sounds [...] I think they're adorable" and "sounds cheery [...] it's a cute little chime." Several participants did notice the difference in sounds from prior sounds heard during operator control; two preferred the new sounds (commenting, for example, that "[they] added more character to Dax" or "he was making more noises than usual, which was kind of cool [...] I guess it kind of felt more interactive") and one preferred the old sounds (stating that the new sounds were "more bass-y and sharp"), though noting that both sounds were agreeable.

5 Follow-up Evaluation: Online Dax Assessment

The in-the-wild evaluation results tended to be positive, but the data that could be feasibly collected in that context was relatively small. To augment the in-person data we collected a set of recordings of the Dax robot interactions and conducted a follow-on online video-based study using Prolific, an online research study platform [13]. This effort was approved by the Oregon State University Institutional Review Board under protocol #IRB-2019-0068.

5.1 Methods

This subsection covers our expectations going into the study, information about our sample, video stimulus information, and details on measurement and analysis methods.

Hypotheses: Based on the results of past similar work (e.g., [26,24]) we established the following hypotheses:

- **H1:** Adding positive-valence emotional sound to a robot will lead to higher perceived warmth, competence, purchasing interest, and value of a robot.
- **H2:** Adding negative-valence emotional sound to a robot will lead to lower perceived warmth, but higher perceived competence, purchasing interest, and value of a robot.

One key difference in these hypotheses compared to past similar work (e.g., [4]) was that we expected that negative-valence emotional sound would make a robot seem less socially warm. (In the past, most of our added sound had a positive valence, so this distinction was unnecessary.)

Fig. 2. Cropped frames from the online survey-based study stimulus videos. From left to right: the "delivery" stimulus, where the robot approaches, stops, and then opens its compartment; "passing", where the robot passes by, tilts its head, and makes a happy facial expression; and "stopping", where the perspective walks in front of the robot, blocking it and causing it to shake its head and make an unhappy facial expression.

Participants: Our sample ($n = 75$) included adults between 19 and 75 years of age ($M = 37.1$, $SD = 14.4$), with 37.3% men (including 1.3% transgender men), 60.0% women, 1.3% nonbinary individuals, and 1.3% agender individuals. Participants included in this sample all passed the study's manipulation check, as further described below.

Study and Stimulus Design: Our sound design for Dax augmented the current set of robot sounds, so our key comparison of interest was how much the newly designed sounds (added sounds) improved the system compared to a base interaction without added sound (original sound). Accordingly, we manipulated sound presence and sought to identify differences in viewer perception without vs. with these added sounds, using a within-subjects design.

The study employed six videos of Dax as stimuli:

1. Original Sound-Delivery: Dax completing a mock delivery with no added sounds.
2. Original Sound-Passing: Dax passing by the viewer, looking at them and making a happy facial expression.
3. Original Sound-Stopped: Dax being stopped by the viewer, shaking its head and making a sad facial expression.
4. Added Sound-Delivery: Dax completing a mock delivery with the added sounds as described above ("arrival," "yes," "ta-da," and "yay").
5. Added Sound-Passing: Dax passing by the viewer, looking at them, making a happy facial expression, and playing "arrival."
6. Added Sound-Stopped: Dax being stopped by the viewer, shaking its head, making a sad facial expression, and playing "downward-no."

Figure 2 shows cropped frames from the video stimuli. All video stimuli used in this study are available in [20].

Procedure: Using these videos, we developed 20-minute online survey. After providing informed consent, participants first completed an introductory module to calibrate their audio device volume. Participants then viewed each of the video stimuli in a counterbalanced order (such that neither the same behavior nor the same sound condition would play consecutively) and answered questions after each video. In the final part of the experiment, participants completed a manipulation check, free-response question, and demographic questionnaire. Participants were compensated with 5 USD for completing the survey.

Measurement: The 20-minute survey included the following measurements:

– After each stimulus, the Robotic Social Attributes Scale (RoSAS) captured participant perceptions of *warmth*, *competence*, and *discomfort* subscales by combining six component attributes for each subscale [2]. Participants rated each attribute on a six-point bipolar Likert scale from "definitely not associated" to "definitely associated."
– After the "Delivery" stimulus, participants also completed the price sensitivity meter (PSM) with wording adjusted to "delivery fee" rather than "product." The PSM is further discussed in our past work [24].
– After all stimuli had been presented, participants provided their thoughts about what influenced their responses most using a free-response question.
– Lastly, participants provided their demographic information.

Analysis: RoSAS results were analyzed using repeated-measures analysis of variance models (rANOVAs) with a significance level of $\alpha = 0.05$, and pairwise comparisons were run on significant rANOVAs with Holm-Bonferroni corrections. We use generalized eta squared for effect size. PSM results were analyzed using Wilcoxon matched-pairs signed-rank tests with the same significance level and Holm-Bonferroni corrections. If significant, extended PSM analysis were run.

5.2 Results

A rANOVA on warmth with just the positive-valence sounds ("Delivery" and "Passing" stimuli) showed a significant increase in warmth ($F(1.00,\ 74.20) = 11.24$, $p = 0.001$, $\eta^2_G = 0.009$) due to sound. However, a rANOVA on warmth with negative valence added sounds ("Stopped" stimulus) did not yield a significant difference. Negative-valence sounds still tended to increase perceived warmth, contributing to an overall increased warmth in rANOVA results for all sounds ($F(1.00, 74.00) = 15.79$, $p < 0.001$, $\eta^2_G = 0.006$) due to sound.

rANOVAs with all sounds showed significant differences in:

– Perceived competence ($F(1.00,\ 74.00) = 4.88$, $p = 0.030$, $\eta^2_G = 0.002$)
– Purchasing interest ($F(1.00,\ 74.00) = 4.36$, $p = 0.040$, $\eta^2_G = 0.003$)

For both competence and purchasing interest, the added-sound condition was rated higher than the original-sound condition. Figure 3 includes a visualization of these responses.

rANOVAs showed no significant differences in perceived discomfort.

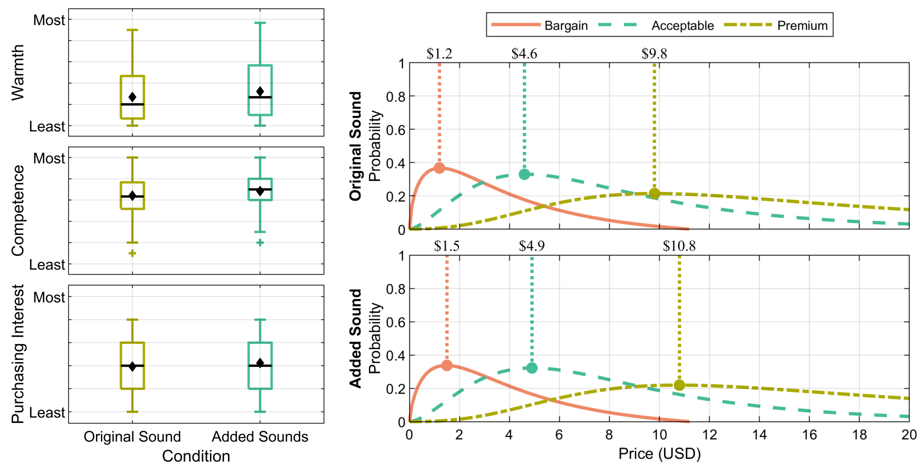

Fig. 3. Results visualizations for the RoSAS and purchasing interest question (left) and PSM analysis (right). The boxplot center lines represent the median, lower and upper lines represent the 25th and 75th percentiles (respectively), whiskers extend to up to 1.5 times the interquartile range, and diamonds indicate the mean. The PSM plots show the typical fitted curves for different price points (bargain, acceptable, and premium) for the original (upper) vs. added (lower) sound conditions.

Holm-Bonferroni-corrected Wilcoxon results for the PSM also showed no significant differences for this inventory. Figure 3 shows the distributions resulting from the PSM analysis, which reveal that pricing for all added-sound curves tend to be higher than for original-sound curves.

6 Discussion

Anecdotes from the in-the-wild evaluation suggested that the design process for adding robot sound to the Dax deliveries was successful, but we sought complementary (and more conclusive) evidence about the design products using the online follow-up evaluation as well. In the follow-up experiment, **H1** and **H2** were partially supported. While negative-valence sounds did not decrease warmth as expected, positive-valence sounds did increase warmth, and all added sounds increased perceived competence and purchasing interest. PSM results did not yield any significant differences, although trends in the PSM results show that added-sound behavior tended to be perceived as more valuable in all typical price categories. Overall, the follow-up results supported the effectiveness of the designed sounds; they usually increased perceived warmth, in addition to generally increasing competence and purchasing interest. Accordingly, this study helped to not only highlight the effectiveness of the sounds themselves, but also to support the effectiveness of the sound design process laid out in our own past efforts, which guided the current investigations.

Strengths of this work included the pursuit of testing and data collection in in-the-wild settings. The inclusion of music technology experts as part of the research team, as well as a robotics company as a key partner, were essential to the success of the work. By the end of the project, we had multiple types of evidence (from anecdotes to statistically significant findings) that supported the idea that our sound design efforts were successful. It is currently unusual to find all of these attributes, each of which strengthen the clout of our efforts, within a single HRI investigation.

At the same time, our work is not without *limitations*. For example, the sample size from the in-the-wild efforts was fairly small, and the methods of interviewing users could (for example) introduce please-the-experimenter bias. Although we did our best to maintain equipoise during the interviews, future similar work could be strengthened by larger samples and multi-method evaluations. On the other hand, experiments conducted in an online setting limit the ecological validity of the results, and it is difficult to control participant experience for uniformity across the group. We used built-in audio adjustment steps and manipulation checks to ensure the data quality as much as possible, but we acknowledge that in-person replication is often needed to solidify the clout of online study results. Additionally, it is important to expand this type of work in the future. The current investigation covered only one robot and one set of sounds, but similar efforts involving additional robots and types of sounds would be needed to understand the broader generalizability of our findings.

7 Conclusions

Fundamentally, design is an open-ended process with unlimited potential processes and products. As such, creating nonverbal sound for HRI may feel daunting to roboticists who have little expertise in sound design. Collaborations can help fill this skill gap, but unguided collaborations may not lead to useful products. In this paper, we used the design thinking prescribed in our own past work as a template for robot-sound-specific guidance; this guidance was inspired in the past by meetings with roboticists and sound designers considering the nonverbal sound question, as well as cross-cultural and cross-disciplinary experiences.

Our author team, which included both expert roboticists and experts in music technology/sound design, put this process to the test, together developing new nonverbal sounds for Dax, a food delivery robot operating locally within Oregon. Through a combination of in-the-wild interviews and online surveys, we showed that the process did create sounds that positively impacted people's perceptions of Dax. For roboticists considering adding nonverbal sound to their robots, we recommend following the design thinking process and the role-specific guidance provided in our past work [22], as we also did in the present work. We believe that this process can help scaffold and hasten the improvement of nonverbal behaviors for day-to-day robots, an important directive for ensuring that these systems can become acceptable and beneficial parts of daily life.

8 Acknowledgments

We thank Daxbot for their willingness to collaborate with us in the efforts described in this paper, particularly Jason Richards, Mason Reeves, and Ryan Chesla.

References

1. Vocable. Wikipedia (Oct 2021), https://en.wikipedia.org/w/index.php?title=Vocable&oldid=1052353406
2. Carpinella, C.M., Wyman, A.B., Perez, M.A., Stroessner, S.J.: The Robotic Social Attributes Scale (RoSAS): Development and Validation. In: Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI). pp. 254–262 (2017)
3. Cha, E., Fitter, N.T., Kim, Y., Fong, T., Matarić, M.J.: Effects of robot sound on auditory localization in human-robot collaboration. In: Proc. of the ACM/IEEE International Conference on Human-Robot Interaction. pp. 434–442 (2018)
4. Chan, L., Zhang, B.J., Fitter, N.T.: Designing and validating expressive cozmo behaviors for accurately conveying emotions. In: Proc. of the IEEE International Conference on Robot & Human Interactive Communication (RO-MAN). pp. 1037–1044 (2021)
5. Dam, R.F., Siang, T.Y.: What is design thinking and why is it so popular? (2022), https://www.interaction-design.org/literature/article/what-is-design-thinking-and-why-is-it-so-popular
6. Daxbot: Daxbot (2024), https://daxbot.com/
7. Langeveld, L., van Egmond, R., Jansen, R., Özcan, E.: Product sound design: Intentional and consequential sounds. Advances in Industrial Design Engineering **47**(3), 47–73 (2013)
8. McGinn, C., Bourke, E., Murtagh, A., Donovan, C., Lynch, P., Cullinan, M.F., Kelly, K.: Meet Stevie: A socially assistive robot developed through application of a 'design-thinking' approach. Journal of Intelligent & Robotic Systems **98**(1), 39–58 (2020)
9. Moore, D., Tennent, H., Martelaro, N., Ju, W.: Making noise intentional: A study of servo sound perception. In: Proc. of the ACM/IEEE International Conference on Human-Robot Interaction. pp. 12–21 (2017)
10. Moreno, P., Nunes, R., Figueiredo, R., Ferreira, R., Bernardino, A., Santos-Victor, J., Beira, R., Vargas, L., Aragão, D., Aragão, M.: Vizzy: A humanoid on wheels for assistive robotics. In: Robot 2015: Second Iberian Robotics Conference: Advances in Robotics, Volume 1. pp. 17–28. Springer (2016)
11. Nwagwu, N., Schneider, A., Syed, I., Zhang, B.J., Fitter, N.T.: The benefits of sound resound: An in-person replication of the ability of character-like robot sound to improve perceived social warmth. Accepted to the 2024 Robotics: Science and Systems Conference (2024)
12. Pelikan, H.R., Jung, M.F.: Designing robot sound-in-interaction: The case of autonomous public transport shuttle buses. In: Proc. of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. pp. 172–182 (2023)
13. Prolific: Prolific (2024), https://www.prolific.com
14. Read, R., Belpaeme, T.: Situational context directs how people affectively interpret robotic non-linguistic utterances. In: Proc. of the ACM/IEEE International Conference on Human-Robot Interaction. pp. 41–48 (2014)
15. Robinson, F.A., Velonaki, M., Bown, O.: Smooth operator: Tuning robot perception through artificial movement sound. In: Proc. of the ACM/IEEE International Conference on Human-Robot Interaction. pp. 53–62 (2021)
16. Shapira Lots, I., Stone, L.: Perception of musical consonance and dissonance: An outcome of neural synchronization. Journal of the Royal Society, Interface **5**(29), 1429–1434 (Dec 2008)

17. Simão, H., Bernardino, A.: User centered design of an augmented reality gaming platform for active aging in elderly institutions. In: icSPORTS. pp. 151–162 (2017)
18. Trovato, G., Paredes, R., Balvin, J., Cuellar, F., Thomsen, N.B., Bech, S., Tan, Z.H.: The sound or silence: Investigating the influence of robot noise on proxemics. In: Proc. of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 713–718 (2018)
19. Wikipedia: Ableton Live — Wikipedia, the free encyclopedia. http://en.wikipedia.org/w/index.php?title=Ableton%20Live&oldid=1226925168 (2024), [Online; accessed 24-June-2024]
20. Zhang, B.J.: Dissertation Appendix – Brian Zhang (2023), https://www.brianzhang.org/publications/dissertation-appendix/
21. Zhang, B.J., Fitter, N.T.: Nonverbal sound in human-robot interaction: A systematic review. ACM Transactions on Human-Robot Interaction **12**(4), 1–46 (2023)
22. Zhang, B.J., Orthmann, B., Torre, I., Bresin, R., Fick, J., Leite, I., Fitter, N.T.: Hearing it out: Guiding robot sound design through design thinking. In: Proc. of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pp. 2064–2071 (2023)
23. Zhang, B.J., Peterson, K., Sanchez, C.A., Fitter, N.T.: Exploring consequential robot sound: Should we make robots quiet and kawaii-et? In: Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 3056–3062 (2021)
24. Zhang, B.J., Sanchez, C.A., Fitter, N.T.: Using the price sensitivity meter to measure the value of transformative robot sound. In: Proc. of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pp. 301–307 (2022)
25. Zhang, B.J., Sigafoos, N., Moffit, R., Syed, I., Adams, L.S., Fick, J., Fitter, N.T.: Sonifyit: Towards transformative sound for all robots. IEEE Robotics and Automation Letters **7**(4), 10566–10572 (2022)
26. Zhang, B.J., Stargu, N., Brimhall, S., Chan, L., Fick, J., Fitter, N.T.: Bringing walle out of the silver screen: Understanding how transformative robot sound affects human perception. In: Proc. of the IEEE International Conference on Robotics and Automation (ICRA). pp. 3801–3807 (2021)