Good Vibrations: How Consequential Sounds Affect Perception of Robotic Arms

Conference Paper · August 2017

Hamish Tennent
Cornell University

Malte Jung
Cornell University

Wendy Ju
Stanford University

2017 26th IEEE International Symposium on
Robot and Human Interactive Communication (RO-MAN)
Lisbon, Portugal, Aug 28 - Sept 1, 2017.

Good Vibrations: How Consequential Sounds Affect Perception of Robotic Arms

Hamish Tennent[1], Dylan Moore[2], Malte Jung[3], & Wendy Ju[4]

Abstract— **How does a robot's sound shape our perception of it? We overlaid sound from high-end and low-end robot arms on videos of the high-end KUKA youBot desktop robotic arm moving a small block in functional (working in isolation) and social (interacting with a human) contexts. The low-end audio was sourced from an inexpensive OWI arm. Crowdsourced participants watched one video each and rated the robot along dimensions of competence, trust, aesthetic, and human-likeness.**

We found that the presence and quality of sound shapes subjective perception of the KUKA arm. The presence of *any* sound reduced human-likeness and aesthetic ratings, however the high-end sound rated better in the competence evaluation in the social context measures when compared to no sound. Overall, the social context increased the perceived competence, trust, aesthetic and human-likeness of the robot. Based on motor sound's significant mixed impact on visual perception of robots, we discuss implications for sound design of interactive systems.

I. INTRODUCTION

Recent technological advances have opened the floodgates for a wave of commercial and domestic robots to enter the public domain. While the notion of working alongside robots started to enter the public's consciousness in the movies of the 1970's and 1980's with classic characters such as R2D2, the possibility continues to permeate popular culture to this day. From products such as the newly released toy Cozmo [34] to the luggage assistance robot arms in New York City hotels [7], the prospect of human-robot interaction for everyday consumers feels less like science fiction and more like science fact.

One often overlooked area of this interaction is generated sound. Demonstration videos of actual robots and interactive systems often mute the actual sounds generated by the robots, and thus mislead how we might feel working next to them. An extreme example of this is the US Army's canceled order of the Boston Dynamics LS3 "Big Dog" military robot because "a loud robot...was going to give away their position." [45]

To perfectly tune interactions to all senses, designers must confront the issue of perceptual sound design in addition to physical form and function. As actuated machines are manufactured and expected to function in close proximity to people in their everyday lives, the noise they create becomes an increasing concern. Langeveld et al. distinguish the *intentional* sound a product makes, such as a tea kettle's whistle, from the *consequential* and *unintentional* sound such a refrigerator's quiet hum [26]. Some have studied the intentional sounds robots use to communicate, such as "non-linguistic utterances" [21], [42], but few understand what is communicated by the consequential sounds of motors and actuators in machines such as automatic doors or vacuum cleaners.

Most machines and robot systems inevitably make consequential sound from motors and gears, and we believe these sounds implicitly and explicitly convey meaning. We explore specifically how this consequential sound colors evaluations of a robot's competence, trust, aesthetics, and human-likeness. Do the sounds captured from high-end and low-end arms respectively raise and lower subjective visual evaluations of the same arm? We also consider two different visual contexts could affect that evaluation: a functional context, where a robot completes a task in isolation, and a social context, where a robot completes a task while interacting with a human. Does the presence of a human moderate the effect of sound?

To answer these questions, we bring together key parameters of human-robot interaction and key methods from product sound design, in which the sound generated by devices has been shown to alter the perceived qualities of the objects. We then describe the methodology behind our comparison of two desktop robot arms at the opposite ends of the price and quality spectrum. We conclude with results and implications from this online video study demonstrating sound's mixed impact on visual evaluation of a robotic arm.

II. BACKGROUND

Researchers have explored many explicit and implicit aspects of robotic communication. Much work has been done

[1]Hamish Tennent is with the Department of Information Science at Cornell University, 343 Campus Rd, Ithaca, NY 14853 ht353@cornell.edu
[2]Dylan Moore is with the Center for Design Research at Stanford University, 424 Panama Mall, Stanford, CA 94305, USA djmoore3@stanford.edu
[3]Malte Jung is with the Department of Information Science at Cornell University, 343 Campus Rd, Ithaca, NY 14853 mfj28@cornell.edu
[4]Wendy Ju is with the Center for Design Research at Stanford University, 424 Panama Mall, Stanford, CA 94305, USA wendyju@stanford.edu

in visual and aesthetic design as well as in the expressiveness of behavior and motion (e.g. [29], [20], [25]). Parallel, but independent, research has explored how sounds shape users' interactions with products.

A. Product Sound Design

Product sound design seeks to understand the component parts of sound and how they effect people's perception of a product. Lyon [28], classifies the sounds emitted by consumer products into four main dimensions:

1) *Strength:* loudness of sound
2) *Annoyance:* roughness or noisiness of sound
3) *Amenity:* regularity or pleasantness of sound
4) *Information content:* identification, performance, or condition conveyed by sound

In this study, we focus on potential for impact on the latter three dimensions as strength is highly environmentally dependent and could easily overwhelm the others. It is worth noting that eliminating noise completely is not always desirable, in the case of a silent food mixer that failed on the market because it was perceived not to have any power. [10]

The information conveyed by sound intuitively forges an impression through interaction. Sound can represent "ergonomics, safety, emotions, hedonics, psychoacoustics, and other attributes depending on product's design" [35, p. 11]. Ju [22] discusses the interaction we have with the consequential sound of an automatic sliding door, whose low hum signals to us that it knows we are approaching. These "implicit interactions" as Ju defined them are critical to creating more natural dialogue between people and products.

Sound is an important channel of communication between robots and humans in industrial settings, as human workers are able to monitor robots without taking their visual attention away from their own tasks [44]. Read and Belpaeme found that sound can also communicate emotional content. Children, for example, ascribe emotion to "non-linguistic utterances" of robots such as "clicks, whirrs, and beeps" [41]. When sound design is neglected, interactions can suffer. For instance, researchers report that the motor sounds of Paro, a pet robot designed for those who cannot otherwise take care of live animals, negatively interfered with interactions [21].

When designing for sound, direct user input is critical to exploring how a sound is perceived. Methods such as *sound quality assessment* invite juries of customers to evaluate sounds using pairwise comparison or evaluation along one more dimensions such as noisiness or clarity. User input is critical as interpretation of sound is "based on a relative evaluation of the sound compared to the expected sound of the product, and is often not based on technical merit or even the reality." [39, p. 989] For some, the roaring Harley Davidson motorcycle is a joyful and wonderful sound, while for others it evokes substantial irritation.

Moore et al. explored perceptual differences between sounds emitted by servo motors, and determined that online observers can differentiate such sounds across metrics such as precision, roughness, and annoyance. They also found that users anthropomorphized sounds without prompting, noting some sounds as "professional" or "without personality." The general impression of these sounds was overall quite negative, suggesting that sounds produced by servos may significantly interfere with interactions [31].

B. Key Elements of Human-Robot Interaction

1) Competence and Trust: Competence and *trust* are core psychological principles in measuring evaluation of autonomous systems. Muir gives us an idea of how competence and trust tenuously interplay, positing, "Each individual will have a criterion of competence beyond which a hypothesis of trust will be adopted, and below which a hypothesis of distrust will be adopted." [32] Hancock et al. [17] found that the perceived competence of a robot had greater association with a person's rating of trust than with environmental or human-related factors.

Barber [1] offers a model of trust proposing that a person's subjective ratings of trust for another person is based upon a taxonomy of (1) persistence of natural law, (2) technically competent performance, and (3) fiduciary responsibility. He expands this to say these expectations also cover many aspects of the steady-state human-machine relationship and that as technology becomes more prevalent. However, when there is no prior history to build assessments on, Rempel et al. describe the process of building trust in terms of (1) subjective *predictability* of future events, (2) *dependability* on the current relationship, and (3) *faith* that things will continue as they are.

Developing this trust is key to building relationships between automated systems and humans. Hancock et al. [17] describe how competence and trust are closely intertwined with functional aspects of the robot as well as human characteristics of self-confidence and expertise. [37] Hancock [17] and Parasuraman [36] found that the more people trusted in the automated system the less likely they were to engage and interrupt it as it progresses towards completion of the task. Bergmann et al. [4] found that the use of gestures in a virtual agent significantly increased a person's perception of the agent's competence. Jung et al. [24] found that the use of back channeling behavior, such as a nod or small non-linguistic utterances, increased perceived engagement but decreased perceived competence of a robot. Takayama and Ju [46] found that when robots physically enacted behaviors reflecting satisfaction or disappointment in their own task performance, people's perceived competence and intelligence of the robot increased compared to situations where the robots did not react at all.

Overall, competence and trust are highly volatile factors closely intertwined with a robot's function and aesthetics. They are also far from objective, and subject to subtle manipulation. Barber's [1] and Rempel et al.'s [43] models give us a definition of trust based on the subjective perception of a machines competence that we will use in this paper.

2) Aesthetics: As *aesthetic* qualities are subjective–that is, they do not exist on their own, but "in the eye of the beholder"– and as there is a two-way relationship between aesthetics and perceived function, attempts to define aesthetics in designed interactive systems are muddy. Desmet and Hekkert offer a definition of "Aesthetic Experience," that is, a "product's capacity to delight one or more of our sensory modalities." [12]

Purists, such as Hassenzahl, offer definitions for the aesthetic qualities of a user's experience with a designed object that frame aesthetics in terms of the "non-instrumental quality, forming an important aspect of product appeal and experience." [19] Norman, conversely, argues that affective perceptions of a product's experience–its "emotional design"–are part of its function. [33] We will adopt Hassenzahl's subsequent conciliatory definition which concedes an interplay between aesthetics and usability, but argues that they are distinct and not directly causal. [18]

This means we include impressions such as whether products are "cool," "adorable," or "strong" are classified as aesthetic judgments [38]. Although what constitutes "aesthetics" is an ongoing debate, this definition allows us to address the more affective and subjective perceptions of how an interactive systems are perceived by a person.

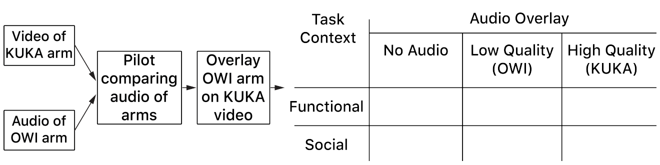

Fig. 1. Timeline showing steps taken through the study.

Fig. 2. The two robot arms we used for our study. On the left is the OWI Robotic Arm Edge and on the right is the KUKA youBot arm. The prices for each arm are $41.50 and $15,700, respectively. For this reason, we define the sounds generated by each arm as low-end and high-end, respectively.

III. Methodology

To explore the impact of auditory and visual stimuli on these core elements of interaction, we formulated a 2x3 (2 contexts x 3 sound conditions) experimental design. We created an online survey asking people to evaluate videos of a Kuka arm moving a small Jenga block around along dimensions of interest.

A. Manipulations

1) Sound Capture and Pilot Testing: We generated three sound conditions to overlay on the videos of a robot arm's movement: *no sound*, *low-end* sound, and *high-end* sound. To generate the low- and high-end sound conditions, we captured sound from two different robot arms. The high-end sound in our study came from a KUKA youBot arm with a retail price of $15,700. The low-end sound came from a consumer grade OWI Robotic Arm Edge with a retail price of $41.50. These two arms can be seen in Figure 2 and represent the upper and lower ends of the consumer focused robot arms on the market today.

In order to verify the subjective differences in the sounds of the two arms we ran a small pilot study on Amazon Mechanical Turk ($N = 24$) asking participants to listen to the pair of sounds and rate them along 6 subjective metrics (strength, precision, pleasantness, expensiveness, roughness, annoyance) as well as overall preference, replicating the methodology that Moore et al. used to examine servo motor sounds [31]. The sounds were recorded using a GoPro camera placed 1 meter away from the robot. The two sound clips were edited so they were both 4.4 seconds long and had peak amplitudes of 16.4db (KUKA arm) and 20.1db (OWI arm). Sounds were presented in a random order with half of participants hearing the OWI arm first and the other half hearing the KUKA arm first. The result of this pairwise comparison was a count indicating how many people preferred one sound over the other along any given dimension, and how strongly they preferred it. These overall counts are shown in Figure 3. Surprisingly, the OWI sound was rated as more precise, however it was also rated as more annoying. There was also a clear overall preference for the KUKA arm.

2) Task Context: The arm moved a Jenga block from a hopper to a specified location. We considered the distinct possibility that a person interacting with a robot would increase empathy for the user or offer other social cues, thus resulting in a measurable effect on how the different sounds would be interpreted. The arm completed this task in two contexts:

1) *Functional*: arm aims to place a block on a preexisting tower of blocks
2) *Social*: arm aims to place a block in a person's hand

We adopt our social context from Breazeal's idea of a sociable robot being one that "proactively engages with humans to satisfy internal social aims" [6], [5], [11]. Dautenhahn states that "social skills are essential for a robot companion. Without these, it will not be accepted on a longterm basis." [11].

B. Video Production

The video and audio were captured using GoPro Hero 3 cameras. The recordings took place in quiet, but not acoustically isolated rooms. For the high-end sound conditions, we used the KUKA arm's natural sound from the video. For the low-end sound conditions, we recorded the OWI arm mimicking similar movements to the KUKA arm and carefully overlaid this sound onto the video. The audio was peak-matched between conditions. In the silent condition, we added the same background noise recorded in the room from the other two videos so that all six had similar background noise.

During the pilot study, subjects in the silent condition commented that they were concerned that the video was not working properly. Because of this, we added a soft door closing sound at the beginning of all videos to subtly hint to participants that the video was working properly.

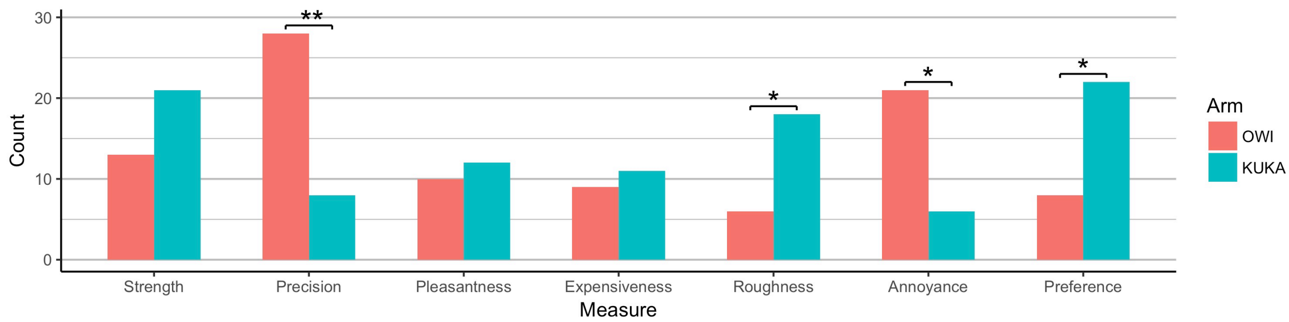

Fig. 3. Pilot test results comparing low- and high-end sounds showed KUKA arm sound was more preferred overall and less annoying. Interestingly, the KUKA arm was rated less precise and more rough than the OWI. Performed χ^2 tests to compare the count with each measure and adjusted p values for multiple comparisons using Benjamini and Hochberg's method [3]. * indicates $p < .05$, ** indicates $p < .01$

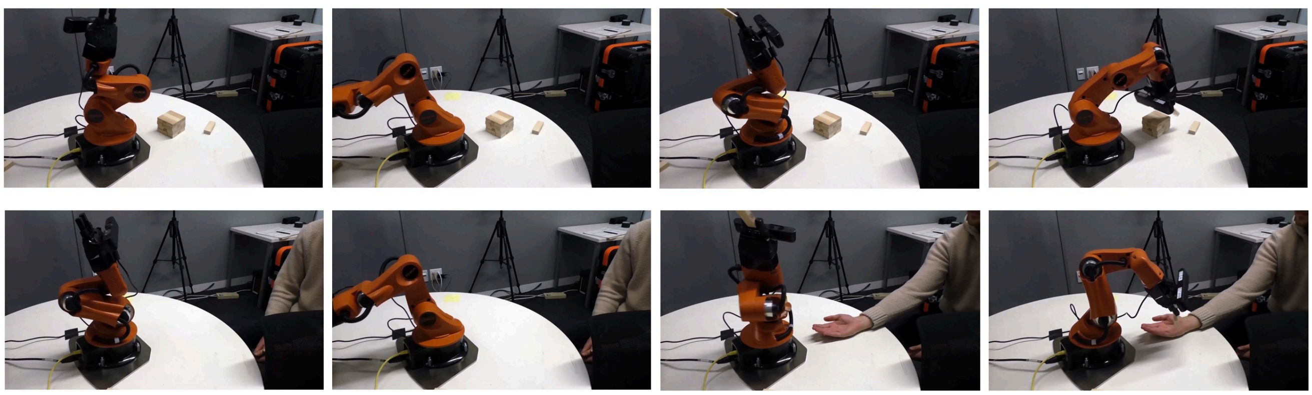

Fig. 4. Top row: Functional condition video frames. Bottom row: Social condition video frames.

We froze the video for two seconds on the last frame, just before the block would have been placed on the existing tower or hand, depending on the context. This was to inject an element of ambiguity into the video on the part of the observer as previously done by Ju in her video study of automatic sliding doors [23]. If the robot successfully completed its task, it may have overshadowed the subtle effect of the sound. After this freeze-frame, the video faded to black to signal the end of the video. Stills of the final videos can be found in Figure 4, and sample videos can be found in the attached video figure.

C. Measures

We asked participants to report their opinion of the robot in the video in distinct question blocks. First, a question asked if the video increased the participant's level of stress on a 5-point scale of "Not at all" to "A great deal." Second, a table asked participants to rate how well a list of nine adjectives, such as "smooth" and "competent," described the arm they saw in the video from "Not well at all" to "Extremely well." Third, a table asked participants to rate their agreement with a list of statements such as "I trusted the actions of the robot." Words and statements with a "(-)" annotation in Table I were reverse coded.

We averaged the responses to these questions into three compound metrics and calculated their consistency using Cronbach's alpha. First, *competence* (α = .82) reflects functional qualities such as precision and reliability. Second, *trust* (α = .80) reflects how comfortable they felt in scenarios where the robot interacted with themselves or another person such as in the workplace or with small children. Third, *aesthetics* (α = .63) reflects the movement and behaviors of the robot through measures such as elegance and calmness. The list of questions and measures can be found in Table I.

1) Competence: We adapted Ever et al.'s [13] 7-point Likert scale questions, modifying the wording from "assistant" to "robot" and from 7 to 5-point. We used this to determine trust in the robot through a two question scale where participants were rating statements from 1 (strongly disagree) to 7 (strongly agree). These statements included "I trusted the actions of the robot" and "the robot was reliable." We also adapted several applicable keywords from the Godspeed Metrics [2]. We expected the robot may be seen as less competent if its sound was less appealing.

2) Trust: We devised several statements such as "I would feel comfortable working alongside this robot" as simple, accessible proxies for measurements of trust. Hancock et al. [17] found a correlation between robot performance factors and the participants' subjective rating of trust in the system. By manipulating performance-related metrics such as perceived precision and strength, we hoped to see a significant correlation with people's development of trust in the system,

TABLE I
INDIVIDUAL WORDS AND PHRASES CONTRIBUTING TO COMPOUND MEASURES.[1]

Competence	Trust	Aesthetics
Competent	Unfriendly (-)	Smooth
Imprecise (-)	"Did this video cause your level of stress or anxiety to increase?"	Calm
Inaccurate (-)	"This robot would be disturbing in my personal office environment." (-)	Elegant
"The robot was reliable."	"I would feel safe leaving this robot in a room with small children."	Inexpensive (-)
"I think the robot was about to successfully place the block."	"I would feel comfortable working alongside the robot."	
	"I trusted the actions of the robot."	

either based on the context or sound condition. In line with Hancock et al.'s finding that human and environment related factors of the interaction do not significantly affect peoples' trust, we expected to see little effect on the subjective rating of trust between the conditions where we do or do not have a person directly interacting with the robot.

3) Aesthetics: Building on the measures used by Moore et al. [31], we included a number of aesthetic measures commonly found in product sound design and human-robot interaction. In product sound design, participants have rated sounds on a number of different criteria, including, but not limited to, acceptability, loudness, power, annoyance, effectiveness, pleasantness, noisiness, gentleness, spaciousness, brightness, and clarity [27], [15], [9]. Because of their inherent subjectivity, we believe aesthetics measures have the best chance of being influenced by sound. Given the negative impressions found my Moore et al., we expected the presence of sound to reduce aesthetic evaluations.

4) Human-Likeness: We adapted a single item from Bartneck et al. [2] asking participants to rate how "humanlike" the robot appeared.

D. Survey Procedure

We created a survey using Qualtrics and placed it onto Amazon Mechanical Turk (AMT) to recruit participants. We chose AMT as a survey platform as it allows easy access to a diverse set of participants. AMT responses have been shown to correlate well with laboratory data and allow for quick data collection and study iterations [30], [8]. While far from definitive, this prototype is valuable to test for predicted effects before conducting expensive and time consuming laboratory studies.

The first page of the survey included brief information and a test video for participants to check that their embedded video player worked correctly. We included explicit instructions asking subjects to use headphones to listen closely to the sound, however one participant reported that they heard the sounds better using their computer's built-in speakers.

The second page included one of the six videos described earlier. Each participant saw only one 22 second video. After watching the video as many times as they wanted, participants answered three blocks of questions described in Section III-C.

The third and last page included questions on demographics, the participant's familiarity with robots and electronics, and a final open-ended feedback box.

E. Participants

Participants (N = 320) were workers on AMT (35% female, .6% transgender, 64.4% male), all from the United States, with previous ratings of 95% or higher and paid $0.50 for the estimated 4 minute task. We estimated this sample size from the statistical power required to see effects across six study conditions. Average time to complete the survey was 3.9 minutes, with a minimum of 38 seconds and a maximum of 15.7 minutes. All participants reported normal hearing without the use of a hearing aid, and reported minimal experience with robotics or electronics.

IV. RESULTS

A. Statistical Approach

For the compound ratings of competence, trust, and aesthetics, component questions were simply averaged. Scores from reverse-coded questions noted in Table I with a "(-)" were flipped before averaging. All ratings have a theoretical range of 1 to 5, with 5 being the positive end, as in more competent or more trustworthy.

Statistical analysis was completed using R version 3.3.2 and RStudio version 1.0.44 [40]. For all measures, the general linear model function `lm()` in the R base package was used to calculate regressions of compound measures and human-likeness from the context and sound conditions. Where homogeneity of variance was not satisfied shown by Levene's test in the `car` package [16], power or inverse data transformations were applied. Compound measure consistency was calculated using `cronbach()` from the `psy` package [14], and figures were generated using `ggplot2` [47]. All reported p values were adjusted using Benjamini and Hochberg's correction for multiple comparisons [3]. Identical corrected p values from different t statistics are a consequence of this method. Error bars in Figures 5 through 9 represent +/- 1 standard error from the mean.

B. Competence

As shown in Figure 5, there was a marginal interaction between context and high-end sound, $b = 3.44,\ t(314) = 2.79,\ p = .087$. In the social context, high-end sound marginally increased competence relative to the no sound condition, $b = 2.33,\ t(314) = 1.97,\ p = .087$. In the functional context, high-end sound slightly decreased competence relative to the bad sound, but not significantly $b = -1.11,\ t(314) = 0.96,\ p = .394$.

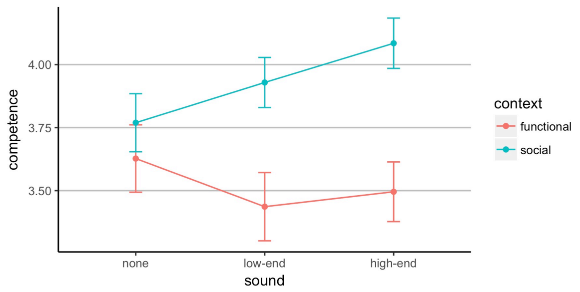

Fig. 5. High-end sound significantly interacted with context, slightly reducing competence evaluation in the functional context, and increasing it in the social context.

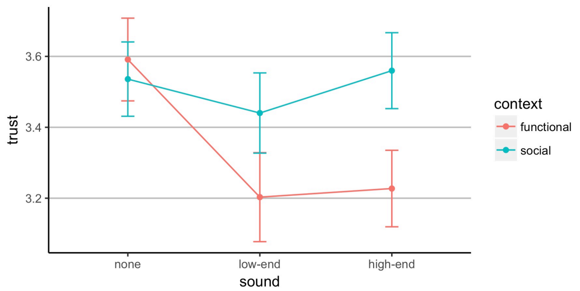

Fig. 6. Participants trusted the robot more in the social context, and the presence of either sound reduced trust in the functional context only.

C. Trust

As shown in Figure 6, social context marginally increased trust overall, $b = 0.17,\ t(316) = 1.89,\ p = .094$. The low-end sound reduced trust, $b = -0.24,\ t(316) = -2.16,\ p = .087$, and the high-end sound had lower trust as well, but the difference from the no sound condition is not significant, $b = -0.17,\ t(316) = -1.53,\ p = .162$. The presence of either sound reduced trust, $b = -0.069,\ t(316) = -2.12,\ p = .087$, but the difference between the sounds themselves is not significant, $b = 0.035,\ t(316) = .631,\ p = .569$.

As expected, trust and competency scores were highly positively correlated, $r(318) = .66,\ p < .001$.

D. Aesthetics

Participants evaluated the robot to be more aesthetically pleasing if it did not play any sound at all, compared to either the low- or high-end sound, $b = -0.844,\ t(316) = -4.37,\ p < .001$. The high-end sound slightly increased aesthetic evaluation relative to the low-end sound, slightly moderating this drop in aesthetic evaluation, but the difference was not significant, $b = 0.53,\ t(314) = 1.63,\ p = .120$.

E. Human-Likeness

As shown in Figure 8, subjects rated the robot arm without sound as more human than either robot with sound, $b = 0.14,\ t(314) = 3.39,\ p = .004$. However the difference between the low- and high-end sounds was not significant, $b = -0.056,\ t(316) = -0.38,\ p = .691$. Overall, subjects rated the robot in the social context as more human than the robot in the functional context, $b = 0.23,\ t(314) = 2.00,\ p = .087$.

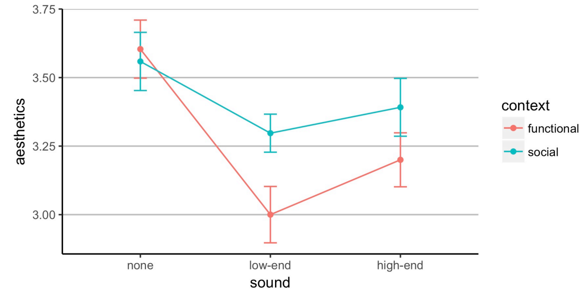

Fig. 7. The presence of either sound (high- or low-end) reduced the aesthetic rating of the arm. In some components of the aesthetic measure, the low-end sound performed worse aesthetically than the high-end sound.

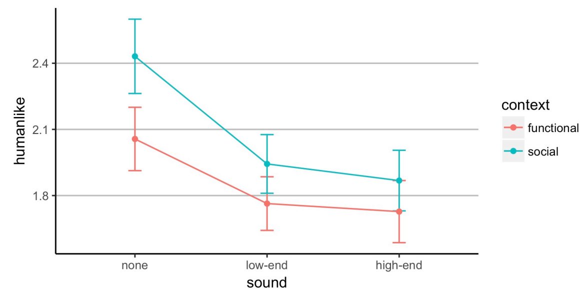

Fig. 8. Human-likeness rating of the robot was significantly lower in the two sound conditions than the no sound condition. The social context marginally increased human-likeness evaluation.

V. Discussion

In general, the presence of sound negatively colored how people perceived a robot in the areas of trust, aesthetics, and human-likeness. However, when compared to a silent robot, high-end sound was able to increase perceived competence when the robot interacted with a human, but decrease perceived competence when the robot was not interacting with a human. As such, context appears to play a significant role in how the sound is interpreted – the robot in the social context had higher ratings nearly across the board. This is likely due to subjects' perceptions of a robot that is able to safely interact with a human, and/or subtle social cues from the actor's behavior in the video. To a certain extent, the social priming seems to moderate the sound's negative effect, especially in the evaluation of trust.

We can offer a potential hypothesis to explain sound's interaction with context: people may perceive an interactive robot as being more competent than one that is operating in isolation, regardless of the associated sound. Sound is generally expected from a robot, so that expectation, coupled with the priming of a robot's competency with human interaction, likely improved evaluation. This interaction speaks to both the complexity and importance of user studies in honing the impressions designers try to elicit.

Sound's overall negative coloring is consistent with Moore et al.'s hypothesis that motor sounds detract from aesthetic experience. [31] However, it is interesting that while participants expressed a significant preference for the KUKA arm sound over the OWI sound, that preference did not translate to significant differences between the two low- and high-end sound conditions. It is possible that the other aspects that differed between the two sounds, such as precision and roughness, confounded this differentiation. Our initial framing of aesthetics was that subjective judgments are independent of functional aspects of the robot or interaction. Within this measure, we attempted to capture an overall visual impression, but did not fully succeed at capturing a difference between the high- and low-end sounds. However, the question still remains: can elements of consequential sound can be designed, where it cannot be eliminated, to mitigate the reduction in evaluation expressed by subjects?

The effect of no sound increasing human likeness aligns with intuition, as most humans do not make grinding gear noises noises as they move. As anthropomorphic and non-anthropomorphic robots become more commonplace, these sounds will be a factor that designers ought to contend with. Eliminating sound may ease comfort for a robot that one desires to be human-like. However, in a social context, it appears that some sound, perhaps like the high-end sound used in our study, can increase evaluations of competence. This is an optimization problem worthy of exploration.

Finally, from the commentary we received from the study, it appears that the sounds used in this study, even the high-end sound, are not very pleasant to listen to. It seems there is much room to improve the quality and characteristics of the sounds robots make. We cannot make claims as to what the characteristics of that sound would be yet, but hope that future research explores how the design of sound could better mitigate its negative impact on evaluations of a robot arm.

VI. Limitations

First, these exploratory data have many potential sources of noise. Participants on AMT do not produce the highest quality responses, as they are often concerned with efficiency more than accurate completion of the task. However, our primary aim was to test for an effect before diving deeper into controlled laboratory studies.

Second, while the sound in the high-end condition showed as preferred over the other our the pilot study, the pilot also showed no significant difference across several other metrics. This may explain why there are not many significant differences between the sound conditions in the videos, only largely main effects of the presence of any sound influencing evaluation. Future studies with interactive systems, and varying specific dimensions of the sounds, will likely better differentiate between the sounds and their subjective effects on perceptions of robots.

VII. Conclusion

This study provides early concrete data suggesting that robot motor sounds negatively color visual perception of interactions. We believe this to be the first data to suggest this link so directly within the context of Human-Robot Interaction scenarios. We have shown there are key areas such as aesthetic and human-like perceptions where the sounds of the robot have significantly reduced how positively people perceived them. These findings should resonate particularly strongly with designers as a significant number of the newly available robots and interactive devices on the market do apply anthropomorphic behaviors and personalities. Such behaviors and personalities are being influenced by consequential sound. While the demonstrated effect is small, we believe this work may motivate attention to sound as an important channel of expression.

Context played an important role in evaluation of the arm, visually priming people to improve evaluation of the arm overall. We believe this is because robots that are able to interact directly with humans are perceived as more sophisticated than those acting in isolation.

No sound performed better than either sound condition in terms of trust, aesthetics, and human-likeness, however the high-end sound improved competence evaluation compared to no sound in the social context only. It is encouraging that high-end sound is potentially able to positively affect functional perceptions of the robot.

Designers forging social relationships between interactive systems and humans now have another element to consider. Consequential sound is an important element of interaction, and we eagerly await the future development of this, as of yet, untapped channel of communication.

VIII. Acknowledgment

We thank Nikolas Martelaro at the Center for Design Research for his helpful input throughout the study. We also thank Stanford MediaX for financial support for this study. Special thanks to Austin Lawrence at Cornell for assisting in video production.

References

[1] Bernard Barber. 1983. *The logic and limits of trust.* Rutgers University Press. 189 pages.

[2] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81. DOI: http://dx.doi.org/10.1007/s12369-008-0001-3

[3] Y Benjamini and Y Hochberg. 1995. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* 57, 1 (1995), 289–300. DOI:http://dx.doi.org/10.2307/2346101

[4] Kirsten Bergmann, Friederike Eyssel, and Stefan Kopp. 2012. A second chance to make a first impression? how appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7502 LNAI (2012), 126–138. DOI:http://dx.doi.org/10.1007/978-3-642-33197-8-13

[5] Cynthia Breazeal. 2003a. Emotion and sociable humanoid robots. *International Journal of Human Computer Studies* 59, 1-2 (2003), 119–155. DOI:http://dx.doi.org/10.1016/S1071-5819(03)00018-1

[6] Cynthia Breazeal. 2003b. Toward sociable robots. *Robotics and Autonomous Systems* 42, 3-4 (2003), 167–175. DOI:http://dx.doi.org/10.1016/S0921-8890(02)00373-1

[7] Althea Chang. 2015. NY midtown robots allow for conversation-free hotel service. (2015). http://cnb.cx/Yobot

[8] Kuan-Ta Chen, Chen-Chi Wu, Yu-Chun Chang, and Chin-Laung Lei. 2009. Crowdsourceable QoE evalutation framework for multimedia content. In *MM*. Beijing, China, 491–500. DOI:http://dx.doi.org/10.1145/1631272.1631339

[9] Sylvain Choisel and Florian Wickelmaier. 2007. Evaluation of multichannel reproduced sound: scaling auditory attributes underlying listener preference. *The Journal of the Acoustical Society of America* 121, 1 (2007), 388–400. DOI:http://dx.doi.org/10.1121/1.2385043
[10] Donald F Cox. 1967. The sorting rule model of the consumer product evaluation process. *Risk taking and information handling in consumer behavior* (1967), 324–369.
[11] Kerstin Dautenhahn. 2007. Socially intelligent robots: dimensions of human-robot interaction. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 362, 1480 (2007), 679–704. DOI:http://dx.doi.org/10.1098/rstb.2006.2004
[12] Pieter Desmet and Paul Hekkert. 2007. Framework of product experience. *International journal of design* 1, 1 (2007).
[13] Vanessa Evers, Talia Brodecki, Pamela Hinds, and Heidy Maldonado. 2008. Relational vs. group self-construal: Untangling the role of national culture in HRI. *3rd ACM/IEEE International Conference on Human-Robot Interaction* (2008), 255–262. DOI:http://dx.doi.org/10.1145/1349822.1349856
[14] Bruno Falissard. 2012. *psy: Various procedures used in psychometry*. https://CRAN.R-project.org/package=psy R package version 1.1.
[15] Anna Fenko, Hendrik N J Schifferstein, and Paul Hekkert. 2011. Noisy products: Does appearance matter? *International Journal of Design* 5, 3 (2011), 77–87.
[16] John Fox and Sanford Weisberg. 2011. *An R Companion to Applied Regression* (second ed.). Sage, Thousand Oaks CA. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion
[17] Peter a. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53, 5 (2011), 517–527. DOI:http://dx.doi.org/10.1177/0018720811417254
[18] Marc Hassenzahl. 2008. *Aesthetics in interactive products: Correlates and consequences of beauty*. Elsevier Ltd. 287–302 pages. DOI:http://dx.doi.org/10.1016/B978-0-08-045089-6.50014-9
[19] Marc Hassenzahl and Noam Tractinsky. 2006. User experienceÂă-Âăa research agenda. *Behaviour & Information Technology* 25, 2 (2006), 91–97. DOI:http://dx.doi.org/10.1080/01449290500330331
[20] Guy Hoffman and Wendy Ju. 2014. Designing Robots With Movement in Mind. *Journal of Human-Robot Interaction* 3, 1 (2014), 89. DOI:http://dx.doi.org/10.5898/JHRI.3.1.Hoffman
[21] Kaoru Inoue, Kazuyoshi Wada, and Yuko Ito. 2008. Effective application of paro: Seal type robots for disabled people in according to ideas of occupational therapists. In *International Conference on Computers for Handicapped Persons*. Springer Berlin Heidelberg. DOI:http://dx.doi.org/10.1007/978-3-540-70540-6{_}197
[22] Wendy Ju. 2015. The Design of Implicit Interactions. *Synthesis Lectures on Human-Centered Informatics* 8, 2 (2015), 1–93. DOI:http://dx.doi.org/10.2200/S00619ED1V01Y201412HCI028
[23] Wendy Ju and Leila Takayama. 2009. Approachability: How people interpret automatic door movement as gesture. *International Journal of Design* 3, 2 (2009), 1–10.
[24] Malte F. Jung, Jin Joo Lee, Nick DePalma, Sigurdur O. Adalgeirsson, Pamela J. Hinds, and Cynthia Breazeal. 2013. Engaging robots: easing complex human-robot teamwork using backchanneling. ACM, 1555âĂŞ1566. http://dl.acm.org/citation.cfm?id=2441954
[25] Heather Knight and Reid Simmons. 2016. Laban head-motions convey robot state: A call for robot body language. *Proceedings - IEEE International Conference on Robotics and Automation* 2016-June (2016), 2881–2888. DOI:http://dx.doi.org/10.1109/ICRA.2016.7487451
[26] Lau Langeveld, René van Egmond, Reinier Jansen, and Elif Özcan. 2013. Product sound design: Intentional and consequential sounds. *Advances in Industrial Design Engineering* (2013). DOI:http://dx.doi.org/10.5772/3415
[27] Richard H. Lyon. 2000. *Designing for Product Sound Quality*. CRC Press.
[28] Richard H. Lyon. 2003. Product sound quality: From perception to design. *Sound and Vibration* 37, 3 (2003), 18–23. DOI:http://dx.doi.org/10.1121/1.4743110
[29] Nikolas Martelaro, Victoria C Nneji, Wendy Ju, and Pamela Hinds. 2016. Tell Me More: Designing HRI to encourage more trust, disclosure, and companionship. In *The Eleventh ACM/IEEE International Conference on Human Robot Interation*. IEEE Press, 181–188.
[30] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods* 44, 1 (2012), 1–23. DOI:http://dx.doi.org/10.3758/s13428-011-0124-6
[31] Dylan Moore, Hamish Tennent, Nikolas Martelaro, and Wendy Ju. 2017. Making Noise Intentional: A Study of Servo Sound Perception. In *Human Robot Interaction*. Vienna, Austria.
[32] Bonnie M Muir. 1987. Trust between humans and machines, and the design of decision aids. *Int. J. Man-Machine Studies* 27 (1987), 527–539. DOI:http://dx.doi.org/10.1016/S0020-7373(87)80013-5
[33] Donald A Norman. 2004. Emotional Design. *Ubiquity* 2004, January (1 2004), 1. DOI:http://dx.doi.org/10.1145/966012.966013
[34] Nathan Olivarez-Giles. 2016. This Cozmo Robot Could Be Your New Best Friend. (2016). https://on.wsj.com/2dQUn5E
[35] Elif Özcan and RenÃľ Van Egmond. 2006. Product sound design and application: Overview. *5th Conference on Design and Emotion 2006* August (2006), 1–19. http://www.scopus.com/inward/record.url?eid=2-s2.0-84865165618&partnerID=40&md5=72df0cdd73269d49815538eb76572b45
[36] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39, 2 (6 1997), 230–253. DOI:http://dx.doi.org/10.1518/001872097778543886
[37] Eui Park, Quaneisha Jenkins, and Xiaochun Jiang. 2008. Measuring Trust of Human Operators in New Generation Rescue Robots. *Proceedings of the JFPS International Symposium on Fluid Power* 2008, 7-2 (2008), 489–492. DOI:http://dx.doi.org/10.5739/isfp.2008.489
[38] Su E. Park, Dongsung Choi, and Jinwoo Kim. 2004. Critical factors for the aesthetic fidelity of web pages: Empirical studies with professional web designers and users. *Interacting with Computers* 16, 2 (2004), 351–376. DOI:http://dx.doi.org/10.1016/j.intcom.2003.07.001
[39] Glenn Pietila and Teik C. Lim. 2012. Intelligent systems approaches to product sound quality evaluations - A review. *Applied Acoustics* 73, 10 (2012), 987–1002. DOI:http://dx.doi.org/10.1016/j.apacoust.2012.04.012
[40] R Development Core Team. 2016. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing Vienna Austria* 0 (2016), 3–900051. DOI:http://dx.doi.org/10.1038/sj.hdy.6800737
[41] Robin Read and Tony Belpaeme. 2012. How to use non-linguistic utterances to convey emotion in child-robot interaction. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 219–220.
[42] Robin Read and Tony Belpaeme. 2016. People Interpret Robotic Non-linguistic Utterances Categorically. *International Journal of Social Robotics* 8, 1 (2016), 31–50. DOI:http://dx.doi.org/10.1007/s12369-015-0304-0
[43] John K. Rempel, John G. Holmes, and Mark P. Zanna. 1985. Trust in close relationships. *Journal of Personality and Social Psychology* 49, 1 (1985), 95–112. DOI:http://dx.doi.org/10.1037/0022-3514.49.1.95
[44] Allison Sauppé and Bilge Mutlu. 2015. The Social Impact of a Robot Co-Worker in Industrial Settings. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15* (2015), 3613–3622. DOI:http://dx.doi.org/10.1145/2702123.2702181
[45] Hope Hodge Seck. 2015. Marine Corps Shelves Futuristic Robo-Mule Due to Noise Concerns. *Military.com* (2015).
[46] Leila Takayama, Doug Dooley, and Wendy Ju. 2011. Expressing thought: improving robot readability with animation principles. *Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2011), 69–76. DOI:http://dx.doi.org/10.1145/1957656.1957674
[47] Hadley Wickham. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Vol. 35. Springer-Verlag, New York. 211 pages. DOI:http://dx.doi.org/10.1007/978-0-387-98141-3