The CANELA Framework: Creating Audio-based Novel Engineered LAnguages for HRI

Author Names Omitted for Anonymous Review. Paper-ID [add your ID here]

***Abstract*—In the Star Wars universe, droids such as R2-D2 and BB-8 speak Droidspeak, a conlang (a constructed, artificial language) that consists of strings of beeps and whistles that resemble sentences. A conlang is made up by linguistic components (rules that structure and give meaning to the sounds, e.g., the beeps and whistles that are strung together to resemble sentences) and can be spoken by someone with a certain voice timbre (the way a voice sounds which is an extra-linguistic component, e.g., R2-D2 sounds slightly different from BB-8). We argue that this framing into separate components of timbre and linguistics is an important one in robot communication design that is currently insufficiently addressed in existing robot sound design frameworks and recommendations. Disentangling timbre from linguistic components enables experimenters and designers to frame their research questions more precisely without being confounded. In this paper, we propose our framework CANELA (Creating Audio-based Novel Engineered LAnguages) that treats robot sound design as the creation of a conlang in which timbre is framed separately from the linguistic components.**

I. Introduction

Previous research has shown the importance of sound design for robot communication. The way a robot speaks can affect people's trust in the robot (e.g., [32, 18]), acceptance (e.g., [33]), likability (e.g., [8, 18]). However, designing robot communication for effective and enjoyable human-robot interaction is not straightforward. People's expectations of a robot's spoken communication change based on the context the robot is in, its appearance, its role, and the user's background [7].

Researchers have previously asked, "What should a robot voice sound like?" However, this question does not cover the whole picture of what is needed to create robot spoken communication, and what is meant by "robot voice" differs across papers; sometimes researchers refer to the timbre of the voice, while others mean the robot's spoken language. Similarly, papers in HRI have introduced frameworks and design recommendations to approach the design of robot voices. While these frameworks and design principles guide designers on specific parts of the design process, to our knowledge, there are no existing frameworks that help design robot sounds from the way they sound, to the creation and selection of a spoken robot language. In our experience, this lack of theoretical background makes it difficult to frame research results that encompass these two different components of robot voice design. For example, people might have preferences for a certain robot language because of the role of the robot but dislike the way it sounds (its timbre) because it does not match the humanlike appearance of the robot. Hence, it makes sense to distinguish robot voice timbre from robot language in robot communication design, although this has rarely been in the foreground of existing design frameworks.

In this paper, we present a framework inspired by basic linguistic principles to guide the design process of a spoken robot language (e.g., Simlish, Droidspeak), which we define as the creation of a "conlang", a language that has been intentionally devised by someone rather than having naturally evolved [24]. The framework distinguishes between linguistic (the way the language and sounds are structured) and extra-linguistic components (timbre, i.e., the way a voice sounds). We identify three main aspects in the framework: robot voice (i.e., timbre), phonetics (the sounds that are possible to make given a voice, e.g., barks for animals or notes for musical robots), and other linguistic properties that make up a language (e.g., semantics, syntax, phonology). We expect that with this framework, researchers will find it easier to tackle specific parts of the design process of robot voice.

II. Design of Robot Communication

A. *Types of robot communication*

There is no clear standardized method for designing robot voices. While there have been some design recommendations [30, 3, 16, 26, 25, 10] and frameworks [40] presented, these are sometimes too specific for the robot's functionalities, part of the design process, or type of utterances. Moreover, these frameworks do not sufficiently address the difference between voice timbre and linguistic properties of a language. This leads to this research not fitting the designer's needs and requirements when wanting to create sounds.

Several types of robot languages have been studied in previous studies. In [17, 21], researchers were inspired by **animal sounds**, while others have used **musical utterances** [15, 32, 31] in designing robot communication. In more recent years, the use of **implicit** [29] or **consequential sounds** [36, 44, 11] (i.e., the unintentional sounds a product makes, such as a refrigerator's hum or sounds as a result of machines' motor movements) for robots have been increasingly investigated. Moving more towards synthetic sounds, so-called **non-linguistic utterances** such as R2D2-like sounds or digital beeps and whirrs, have been used as well [34, 13, 27, 28, 4] where sometimes the sounds are created with help from sound designers/experts, e.g., [13]. Another type of robot communication used is that of **gibberish** [41, 42, 5], which consist of language-like utterances with no semantic meaning (e.g., Simlish). Researchers have also studied how adding **paralinguistic utterances**, such as laughter or "argh", can enhance HRI [14, 18], although they have rarely been studied in isolation. Similarly, musical utterances and non-linguistic utterances have also been studied in combination with natural language, with the results of this combination being sometimes positively [28] and sometimes negatively [4] perceived by participants. Finally, another option for human-robot communication is to use **speech (spoken natural language)** but designed with different acoustic properties to make robots sound more natural ("humanlike") or synthetic ("robotlike"). Although the research on synthetic/robotlike voices is limited, researchers have used filters and various sound manipulation techniques [39] to turn humanlike voices into more "robot-sounding" ones. These efforts refer more to voice timbre-like aspects of robot communication.

As exemplified by the varied types of languages and voices that researchers have been investigating, we can conclude that 1) robot language and voice timbre are rarely identified as separate elements that need to be studied together to design robot communication, and 2) the design space for robot communication is large. Only recently, efforts have been made into identifying potential factors that can guide robot communication design.

B. Existing frameworks guiding robot communication design

In HRI literature, we find several taxonomies [43] and sound design frameworks that intend to help designers frame and design robot communication. Robinson et al. [30] introduces nine design principles for creating robot sounds. These aim to provide generalizable design recommendations for robot sound design inspired by existing frameworks on film sound and product sound designers while adding insights from commercial robot sound designers. Cambre and Kulkarni's framework [6] aims to guide the design of smart voices, highlighting the importance of the complex interplay between the characteristics of user, device, and context and voice. Seaborn et al. [35] offer a high-level classification framework that describes key factors in voice-in-human-agent-interaction design. In [25], a 3-layered sonification framework is presented that maps robot features to audio parameters. Based on a workshop held with robotics experts, 3 layers of robot-related information that could be displayed via sound were identified: 1) physical base (e.g., size, interactivity), 2) internal state (e.g., urgency), 3) action (e.g., speed).

However, we argue that the fundamental question that is at the basis of each of the design frameworks discussed above is only partially addressed in these frameworks, namely *how should a robot communicate*?, in terms of the type of language it should use and how it should sound like (timbre). While some research has explicitly focused on how a robot should sound by referring to its timbre [38, 18], we also observe that sometimes in literature, it is not clear whether "robot voice design" or "sound design for robots" refers to the robot's timbre (i.e., voice identity) or the robot's language (e.g., gibberish), or both (confounding). This creates a gap in the current design frameworks and guidelines that makes the reporting of results that tackle these two components of robot communication difficult to report on.

C. What factors should be taken into account when designing robot communication?

There are several considerations to take into account when designing a robot's voice, as these and their alignment with people's expectations can influence their perception of the robot and its voice. Research has shown that the perception of a robot is affected by the interplay between its appearance, context, function [37], personality [2], and the users it interacts with [6], as well as how well these elements match its voice [37, 9, 22, 12, 6, 19]. Researchers agree that "how an autonomous agent looks, sounds and behaves should align" [22]. In line with the "matching hypothesis", this alignment can improve people's acceptance of and cooperation with the robot [12], as well as decrease the gap between user expectations and system capabilities [19, 22].

Misalignment of several of these aspects may lead to negative effects on user perception. For example, previous work has shown that a misalignment between the humanness of a robot's face and its voice can lead to an uncanny valley effect [20]. Further, it is hypothesized by [22, 23] that when visual, vocal, behavioral, and cognitive affordances of an agent are not aligned, a "habitability gap" causes increased flexibility of the spoken dialogue system leads to a catastrophic drop in usability. These findings highlight the considerations that should be considered when designing spoken robot sounds and how the success of how these are managed affects the user's perception and interaction with the robot.

III. Proposed Framework

The proposed conceptual framework (see Fig. 2) guides robot sound designers on the creation of a spoken robot *conlang*, an engineered language whose linguistic properties are artificially created. The framework consists of linguistic and extra-linguistic components. We will dissect its components and provide questions a designer needs to answer and make decisions on.

A. Motivation of the framework

In our own research on the perception of different auditory robot communication design, we felt the need for a tool or model to frame our results in. In a series of user studies, we showed participants a picture of the Harmony robot, a healthcare delivery robot, and asked for participant's preferences for two different robot voices: gibberish (EMOGIB [42]) and non-linguistic utterances (BEST corpus [13]).

When asked about their reasoning behind their preference, the responses we received were about how the voice of the robot sounds (timbre), the way the sounds were structured into a language or both. Participants wrote "*that sounds more like a robot*" or "*I find the voice a bit annoying because it is somewhat dramatic, high-pitched, and childish.*" which refers to the timbre of the robot voice. While others referred more to language-like aspects: "*The sounds in B sound more like language and therefore fit better with a robot.*", "*Set B resembles words more, and then people will quickly start looking for the linguistic meaning behind it, which seems*

Fig. 1. [redacted] robot by designed by [redacted]

unnecessarily complex to me.” or a mix of both: “*The robot does not resemble a human enough to use a human voice, even though there is no real speech involved.*” These type of responses in combination with the lack of guidelines on how to approach and frame the design of robot communication in terms of timbre and language-like aspects, motivated us to propose the CANELA framework (see Fig. 2).

B. Description of the framework

When designing spoken robot communication, we propose to frame this as designing a spoken robot language, a conlang. Like a natural language, a conlang consists of several linguistic components. Previous work designing robot voices in HRI has also taken a similar approach to sound synthesis. Namely, they also defined robot sounds as conlangs and considered linguistic properties when designing a *language composer* that created sounds like R2-D2's [34].

The main linguistic component in our framework is the *phonetics* component, where the designer has to define the possible sounds the robot can make. In natural language, these are vowels and consonants (i.e., phonemes), while in conlang, these could be a small set of beeps and whirrs. The other linguistic components, e.g., phonology, syntax, semantics etc., then need to be designed to give structure and meaning to these sounds. For example, in natural language, phonological and syntactical rules dictate how vowels and consonants can form syllables, words, and sentences. Likewise, these linguistic rules can also be designed for beeps and whirrs. Although these rules can be very simple or even non-existent in a limited conlang, a decision should be made in this regard. For example, we can design a very simple conlang that consists of 2 beeps and 3 whirrs, each expressing a different emotion. Finally, the designer has to ask themselves how the “speaker” should sound, what *timbre* the “speaker” should have - this is an acoustic, extralinguistic aspect outside the linguistic system. Timbre refers to voice quality or “That attribute of auditory sensation in terms of which a subject can judge that two sounds, similarly presented and having the same loudness and pitch, are dissimilar.” [1]. For example, the designer can choose between a robotic and humanlike-sounding voice. All these choices need to be informed by how well the robot communication, i.e. timbre *and* linguistic system, align with the robot's appearance, context, function, and behaviors.

In sum, our framework CANELA (Creating Audio-based Novel Engineered LAnguages for HRI, see Fig. 2 and Table I) asks designers to make design choices about a) the possible sounds that the robot can make, b) the rules that structure and give meaning to these sounds, and c) the timbre of the voice, that align with the robot's appearance, context, function, and behaviors.

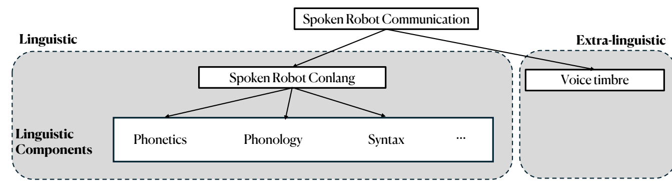

Fig. 2. The CANELA (Creating Audio-based Novel Engineered LAnguages) framework

C. Framework Example: Simple Droidspeak Conlang

We will demonstrate how the framework can guide the design process through an example. In the Star Wars universe, droids such as R2-D2 and BB-8 use Droidspeak to communicate. In this example, we aim to create a voice for a droid that resembles Luke Skywalker's, utilizing a simplified version of Droidspeak.

How do we communicate through sounds? We want to create a simple language inspired by the phonetics of the Droidspeak conlang (the language spoken by some droids in Star Wars, e.g., R2-D2 and BB-8).

What are possible sounds in the design space? While more sounds are possible in Droidspeak, we reduced the possible sounds to the whistles, chirps, and beeps from Droidspeak.

How do you structure the sounds to create a conlang? Within this sound design space, we want to communicate three emotions (happiness, anger, and sadness) and two intentions (greeting and thank you). This is formalized in a dictionary with five sounds that map to these emotions and intentions. We decided to keep the linguistic system of this conlang very simple; there will be no rules regarding syntax, morphology, and phonology or hierarchical structure.

How does the speaker sound? We want the speaker to sound like Luke Skywalker.

This will result in a voice that sounds like Luke Skywalker speaking a simplified version of Droidspeak to communicate three emotions and two intents.

By following our framework, designers can make informed decisions on robot sound design and more precisely identify specific areas for improvement.

[1]First author's cat. A cute and cuddly European shorthair orange/cream cat

TABLE I
THE CANELA FRAMEWORK WITH COMPONENTS, QUESTIONS, AND EXAMPLES EXPLAINED

	Component	Question	Examples
Linguistics	Spoken Robot Conlang	How do we communicate through sounds?	Simlish, R2-D2, Klingon
	Phonetics	What are possible sounds in the design space?	Human vocal sounds, cat sounds, guitar sounds
	Other linguistic properties (e.g., Semantics, morphology, syntax, etc.)	How do you structure and give meaning to the sounds to create a "language"?	Decide on what units to use, e.g., syllables, words; Decide on the dictionary (e.g., intention, emotion)
Extra-linguistic	Voice identity	How does the "speaker" sound?	Timbre, pitch, intensity, speaking rate (e.g., Canela[1], Garfield)

IV. Conclusion

With the CANELA framework, we have presented a way to facilitate the design process by framing the design choices in a) the possible sounds that the robot can make, b) the rules that structure and give meaning to these sounds, and c) the timbre of the voice. It is important that these are informed design choices that align with the robot's appearance, context, function, personality, and its future users. While the presented framework is a first version, it has been proven fruitful in our own research, and we are eager to find out how this framework evolves and can guide future robot communication design challenges.

Acknowledgments

[redacted]

References

[1] ANSI. *Psychoacoustic Terminology: Timbre*. American National Standards Institute, 1960.

[2] Matthew P Aylett, Selina Jeanne Sutton, and Yolanda Vazquez-Alvarez. The right kind of unnatural: designing a robot voice. In *Proceedings of the 1st international conference on conversational user interfaces*, pages 1–2, 2019.

[3] Matthew P. Aylett, Selina Jeanne Sutton, and Yolanda Vazquez-Alvarez. The right kind of unnatural: designing a robot voice. In *Proceedings of the 1st International Conference on Conversational User Interfaces*, pages 1–2, Dublin Ireland, August 2019. ACM. ISBN 978-1-4503-7187-2. doi: 10.1145/3342775.3342806. URL https://dl.acm.org/doi/10.1145/3342775.3342806.

[4] Matthew P Aylett, Yolanda Vazquez-Alvarez, and Skaiste Butkute. Creating robot personality: effects of mixing speech and semantic free utterances. In *Companion of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 110–112, 2020.

[5] Cynthia Breazeal. Toward sociable robots. *Robotics and autonomous systems*, 42(3-4):167–175, 2003.

[6] Julia Cambre and Chinmay Kulkarni. One voice fits all? social implications and research challenges of designing voices for smart devices. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–19, 2019.

[7] Julia Cambre and Chinmay Kulkarni. One Voice Fits All?: Social Implications and Research Challenges of Designing Voices for Smart Devices. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–19, November 2019. ISSN 2573-0142. doi: 10.1145/3359325. URL https://dl.acm.org/doi/10.1145/3359325.

[8] Friederike Eyssel, Dieta Kuchenbrandt, and Simon Bobinger. 'If you sound like me, you must be more human': on the interplay of robot and user features on human-robot acceptance and anthropomorphism. 2012.

[9] Kerstin Fischer and Oliver Niebuhr. Which voice for which robot? designing robot voices that indicate robot size. *ACM Transactions on Human-Robot Interaction*, 12 (4):1–24, 2023.

[10] Kerstin Fischer, Oliver Niebuhr, Rosalyn M Langedijk, and Selina Eisenberger. I Shall Know you by your Voice – Melodic and Physical Dominance in the Design of Robot Voices.

[11] Emma Frid and Roberto Bresin. Perceptual Evaluation of Blended Sonification of Mechanical Robot Sounds Produced by Emotionally Expressive Gestures: Augmenting Consequential Sounds to Improve Non-verbal Robot Communication. *International Journal of Social Robotics*, 14(2):357–372, March 2022. ISSN 1875-4791, 1875-4805. doi: 10.1007/s12369-021-00788-4. URL https://link.springer.com/10.1007/s12369-021-00788-4.

[12] Jennifer Goetz, Sara Kiesler, and Aaron Powers. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003.*, pages 55–60. Ieee, 2003.

[13] Helen Hastie, Pasquale Dente, Dennis Küster, and Arvid Kappas. Sound emblems for affective multimodal output of a robotic tutor: a perception study. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pages 256–260, Tokyo Japan, October 2016. ACM. ISBN 978-1-4503-4556-9. doi: 10.1145/2993148.2993169. URL https://dl.acm.org/doi/10.1145/2993148.2993169.

[14] Koji Inoue, Divesh Lala, and Tatsuya Kawahara. Can a robot laugh with you?: Shared laughter generation for empathetic spoken dialogue. *Frontiers in Robotics and AI*, 9:234, 2022.

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

[16] Jonas Jørgensen and Mads Bering Christiansen. The Sounds of Softness. Designing Sound for Human-Soft Robot Interaction. *Frontiers in Robotics and AI*, 8: 674121, October 2021. ISSN 2296-9144. doi: 10.3389/frobt.2021.674121. URL https://www.frontiersin.org/articles/10.3389/frobt.2021.674121/full.

[17] Beáta Korcsok, Tamás Faragó, Bence Ferdinandy, Ádám Miklósi, Péter Korondi, and Márta Gácsi. Artificial sounds following biological rules: A novel approach for non-verbal communication in HRI. *Scientific Reports*, 10(1):7080, April 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-63504-8. URL https://www.nature.com/articles/s41598-020-63504-8.

[18] Xiaozhen Liu, Jiayuan Dong, and Myounghoon Jeon. Robots' "Woohoo" and "Argh" Can Enhance Users' Emotional and Social Perceptions: An Exploratory Study on Non-lexical Vocalizations and Non-linguistic Sounds. *ACM Transactions on Human-Robot Interaction*, 12(4): 1–20, December 2023. ISSN 2573-9522. doi: 10.1145/3626185. URL https://dl.acm.org/doi/10.1145/3626185.

[19] Ewa Luger and Abigail Sellen. " like having a really bad pa" the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 5286–5297, 2016.

[20] Wade J Mitchell, Kevin A Szerszen, Amy Shirong Lu, Paul W Schermerhorn, Matthias Scheutz, and Karl F MacDorman. A Mismatch in the Human Realism of Face and Voice Produces an Uncanny Valley. *i-Perception*, 2 (1):10–12, January 2011. ISSN 2041-6695, 2041-6695. doi: 10.1068/i0415. URL http://journals.sagepub.com/doi/10.1068/i0415.

[21] Roger K Moore. A real-time parametric general-purpose mammalian vocal synthesiser. In *INTERSPEECH*, pages 2636–2640. San Francisco, CA, 2016.

[22] Roger K Moore. Is spoken language all-or-nothing? implications for future speech-based human-machine interaction. In K Jokinen and G. Wilcock, editors, *Dialogues with Social Robots: Enablements, Analyses, and Evaluation*, pages 281–291. Springer, 2017.

[23] Roger K. Moore. A 'Canny' Approach to Spoken Language Interfaces. In *In CHI-19 Workshop on Mapping Theoretical and Methodological Perspectives for Understanding Speech Interface Interactions*. arXiv, 2019. URL http://arxiv.org/abs/1908.08131. arXiv:1908.08131 [cs].

[24] Arika Okrent. *In the land of invented languages: Esperanto rock stars, Klingon poets, Loglan lovers, and the mad dreamers who tried to build a perfect language*. Random House, 2009.

[25] Bastian Orthmann, Iolanda Leite, Roberto Bresin, and Ilaria Torre. Sounding Robots: Design and Evaluation of Auditory Displays for Unintentional Human-robot Interaction. *ACM Transactions on Human-Robot Interaction*, 12(4):1–26, December 2023. ISSN 2573-9522. doi: 10.1145/3611655. URL https://dl.acm.org/doi/10.1145/3611655.

[26] Hannah Pelikan. *Robot Sound in Interaction : Analyzing and Designing Sound for Human-Robot Coordination*, volume 36 of *Studies in Language and Culture*. Linköping University Electronic Press, Linköping, June 2023. ISBN 9789180751162. doi: 10.3384/9789180751179. URL https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-193560.

[27] Robin Read and Tony Belpaeme. 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 - HRI '12*, page 219, Boston, Massachusetts, USA, 2012. ACM Press. ISBN 978-1-4503-1063-5. doi: 10.1145/2157689.2157764. URL http://dl.acm.org/citation.cfm?doid=2157689.2157764.

[28] Robin Read and Tony Belpaeme. Non-linguistic utterances should be used alongside language, rather than on their own or as a replacement. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 276–277, Bielefeld Germany, March 2014. ACM. ISBN 978-1-4503-2658-2. doi: 10.1145/2559636.2559836. URL https://dl.acm.org/doi/10.1145/2559636.2559836.

[29] Frederic Anthony Robinson, Oliver Bown, and Mari Velonaki. Implicit Communication through Distributed Sound Design: Exploring a New Modality in Human-Robot Interaction. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 597–599, Cambridge United Kingdom, March 2020. ACM. ISBN 978-1-4503-7057-8. doi: 10.1145/3371382.3377431. URL https://dl.acm.org/doi/10.1145/3371382.3377431.

[30] Frederic Anthony Robinson, Oliver Bown, and Mari Velonaki. Designing Sound for Social Robots: Candidate Design Principles. *International Journal of Social Robotics*, 14(6):1507–1525, August 2022. ISSN 1875-4791, 1875-4805. doi: 10.1007/s12369-022-00891-0. URL https://link.springer.com/10.1007/s12369-022-00891-0.

[31] Frederic Anthony Robinson, Mari Velonaki, and Oliver Bown. Crafting the Language of Robotic Agents: A vision for electroacoustic music in human–robot interaction. *Organised Sound*, 27(2): 206–218, August 2022. ISSN 1355-7718, 1469-8153. doi: 10.1017/S1355771822000358. URL https://www.cambridge.org/core/product/identifier/S1355771822000358/type/journal_article.

[32] Richard Savery, Lisa Zahray, and Gil Weinberg. Emotional musical prosody for the enhancement of trust: Audio design for robotic arm communication. *Paladyn, Journal of Behavioral Robotics*, 12(1):454–467, December 2021. ISSN 2081-4836. doi: 10.1515/pjbr-2021-0033. URL https://www.degruyter.com/document/doi/10.1515/pjbr-2021-0033/html.

[33] Simon Schreibelmayr and Martina Mara. Robot Voices in Daily Life: Vocal Human-Likeness and Application Context as Determinants of User Acceptance. *Frontiers in Psychology*, 13:787499, May 2022. ISSN 1664-1078. doi: 10.3389/fpsyg.2022.787499. URL https://www.frontiersin.org/articles/10.3389/fpsyg.2022.787499/full.

[34] Markus Schwenk and Kai O. Arras. R2-D2 Reloaded: A flexible sound synthesis system for sonic human-robot interaction design. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 161–167, Edinburgh, UK, August 2014. IEEE. ISBN 978-1-4799-6765-0 978-1-4799-6763-6. doi: 10.1109/ROMAN.2014.6926247. URL http://ieeexplore.ieee.org/document/6926247/.

[35] Katie Seaborn, Norihisa P. Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. Voice in Human–Agent Interaction: A Survey. *ACM Computing Surveys*, 54(4): 1–43, May 2022. ISSN 0360-0300, 1557-7341. doi: 10.1145/3386867. URL https://dl.acm.org/doi/10.1145/3386867.

[36] Hamish Tennent, Dylan Moore, Malte Jung, and Wendy Ju. Good vibrations: How consequential sounds affect perception of robotic arms. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 928–935, Lisbon, August 2017. IEEE. ISBN 978-1-5386-3518-6. doi: 10.1109/ROMAN.2017.8172414. URL http://ieeexplore.ieee.org/document/8172414/.

[37] Ilaria Torre, Adrian Benigno Latupeirissa, and Conor McGinn. How context shapes the appropriateness of a robot's voice. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 215–222. IEEE, 2020.

[38] Pol van Rijn, Silvan Mertes, Kathrin Janowski, Katharina Weitz, Nori Jacoby, and Elisabeth André. Giving Robots a Voice: Human-in-the-Loop Voice Creation and open-ended Labeling. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–34, May 2024. doi: 10.1145/3613904.3642038. URL http://arxiv.org/abs/2402.05206. arXiv:2402.05206 [cs].

[39] Sarah Wilson and Roger K Moore. Robot, Alien and Cartoon Voices: Implications for Speech-Enabled Systems. In *Proceedings of 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR)*, 2017.

[40] Selma Yilmazyildiz. Semantic-Free Affective Speech Framework for Social Human-Robot Interaction.

[41] Selma Yilmazyildiz, David Henderickx, Bram Vanderborght, Werner Verhelst, Eric Soetens, and Dirk Lefeber. EMOGIB: Emotional Gibberish Speech Database for Affective Human-Robot Interaction. In Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin, editors, *Affective Computing and Intelligent Interaction*, volume 6975, pages 163–172. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-24570-1 978-3-642-24571-8. doi: 10.1007/978-3-642-24571-8_17. URL http://link.springer.com/10.1007/978-3-642-24571-8_17. Series Title: Lecture Notes in Computer Science.

[42] Selma Yilmazyildiz, Werner Verhelst, and Hichem Sahli. Gibberish speech as a tool for the study of affective expressiveness for robotic agents. *Multimedia Tools and Applications*, 74(22):9959–9982, November 2015. ISSN 1380-7501, 1573-7721. doi: 10.1007/s11042-014-2165-1. URL http://link.springer.com/10.1007/s11042-014-2165-1.

[43] Selma Yilmazyildiz, Robin Read, Tony Belpeame, and Werner Verhelst. Review of Semantic-Free Utterances in Social Human–Robot Interaction. *International Journal of Human-Computer Interaction*, 32(1):63–85, January 2016. ISSN 1044-7318, 1532-7590. doi: 10.1080/10447318.2015.1093856. URL http://www.tandfonline.com/doi/full/10.1080/10447318.2015.1093856.

[44] Brian J. Zhang, Knut Peterson, Christopher A. Sanchez, and Naomi T. Fitter. Exploring Consequential Robot Sound: Should We Make Robots Quiet and Kawaii-et? In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3056–3062, September 2021. doi: 10.1109/IROS51168.2021.9636365. URL http://arxiv.org/abs/2104.02191. arXiv:2104.02191 [cs].