International Journal of Social Robotics
https://doi.org/10.1007/s12369-022-00891-0

Designing Sound for Social Robots: Candidate Design Principles

Frederic Anthony Robinson[1] · **Oliver Bown**[2] · **Mari Velonaki**[1]

Accepted: 7 May 2022

Abstract
How can we use sound and music to create rich and engaging human-robot interactions? A growing body of HRI research explores the many ways in which sound affects human-robot interactions and although some studies conclude with tentative design recommendations, there are, to our knowledge, no generalised design recommendations for the robot sound design process. We address this gap by first investigating sound design frameworks in the domains of product sound design and film sound to see whether practices and concepts from these areas contain actionable insights for the creation of robot sound. We then present three case studies, detailed examinations of the sound design of commercial social robots Cozmo and Vector, Jibo, and Kuri, facilitated by expert interviews with the robots' sound designers. Combining insights from the design frameworks and case studies, we propose nine candidate design principles for robot sound which provide (1) a design-oriented perspective on robot sound that may inform future research, and (2) actionable guidelines for designers, engineers and decision-makers aiming to use sound to create richer and more refined human-robot interactions.

Keywords Human-robot interaction · Sonic interaction design · Design principles · Social robots

1 Introduction

Social robots need to exhibit rich and engaging behaviour in order to successfully integrate into human environments [6]. Sound is one of the core modalities robots can use to communicate with humans and a sizeable body of work has explored the many ways in which sound properties influence how humans perceive and interact with robotic agents. When it comes to harnessing the potential of this modality, sound designers find themselves in an application context that shares some similarities with areas such as product, film, or video game sound, but which also comes with many unique challenges and opportunities. Some of the questions designers face include, for example: (1) What robot behaviours should be accompanied by sound and what functions should sound fulfil in these interaction scenarios? (2) What are the implications of designing sound for different robot embodiments with idiosyncratic acoustic characteristics? (3) How does the robot sound design process differ from sound design for linear media, and how should content production practices and techniques adapt? A possible broad categorisation of sound in HRI is shown in Fig. 1. It should be noted, however, that there is no strict separation between categories and some overlap is unavoidable. A comprehensive review of sound in HRI is beyond the scope of this paper.

Sound uttered by robots comprises speech and non-speech utterances. Beyond the semantic content of speech, voice qualities such as human-likeness [3,17,52], whispering [35], speech rate and gender [12], accent [29], and intonation [1,18] can affect HRI in various ways. The effects of voice characteristics are an active area of research in HRI and the wider human-computer interaction (HCI) community, and Cambre and Kulkarni have proposed a set of guiding questions for voice design [9]. For this reason, the scope of this paper does not extend to the voice characteristics of semantic speech. Semantic-free utterances include gibberish speech and musical utterances, among others. A comprehensive review of semantic-free utterances in HRI can be found in [55]. *Sound associated with robot movement* includes consequential sound, the sound of a robot's motors and joints

✉ Frederic Anthony Robinson
frederic.robinson@unsw.edu.au

Oliver Bown
o.bown@unsw.edu.au

Mari Velonaki
mari.velonaki@unsw.edu.au

[1] Creative Robotics Lab, University of New South Wales, Sydney, Australia

[2] Interactive Media Lab, University of New South Wales, Sydney, Australia

Published online: 08 June 2022

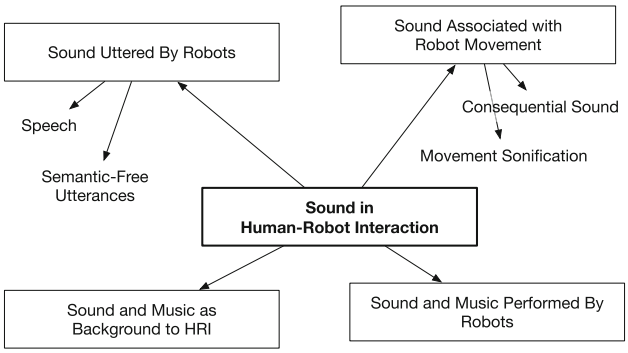

Fig. 1 A high-level categorisation of sound in human-robot interaction

[31,48], and movement sonificiation, artificial sound added to robot movement to influence perception [16,19]. A review of consequential sound and movement sonification can be found in [42]. *Sound and music performed by robots* comprises robotic agents creating music, either on their own [46,56] or in collaboration with humans [13,47]. A review of robotic musicianship by Bretan and Weinberg can be found in [7]. Lastly, *sound and music as background to HRI* includes shared listening experiences [21,57], and more unusual scenarios such as ambient infrasound [49] or surface vibrations that convey a system's confidence level [25]. As demonstrated above, the design space for robot sound is broad and HRI literature does provide some design recommendations for clearly defined interaction scenarios. Beyond sound, there are broad design recommendations for HRI scenarios in general, as, for example, proposed by Tokin et al. [50]. However, there are, to our knowledge, no general guidelines for sound in human-robot interaction which encompass the many diverse challenges and opportunities around designing robot sound. We therefore set out to address this gap by proposing a set of candidate design principles, aimed to provide designers, engineers and decision-makers with comprehensive and actionable guidelines on how to create effective and refined robot sound. We describe these as *candidate* design principles, because (1) they are partly sourced from domains outside of HRI and (2) they are still to be validated through more extensive application within HRI.

In this paper, we briefly survey existing design recommendations on sound in HRI, before introducing a number of design frameworks from the areas of product sound design and film sound. We then present three commercial robot sound design case studies in the form of expert interviews with the robots' sound designers and combine the existing sound design frameworks with insights obtained from the case studies to arrive at candidate design principles. We conclude with a discussion of the current limitations of our design principles.

2 Design Recommendations

This section presents a selection of existing design frameworks. After a brief survey into sound design recommendations in prior HRI work, it looks at HRI scenarios from two perspectives: (1) robots as objects to be sonified, and (2) human-robot interaction as a narrative to be scored. This view allows us to introduce design frameworks from literature in the areas of product sound design and film sound. By then later drawing connections between these design frameworks and the design work performed across the three case studies in Sect. 3, we investigate how they could apply to robot sound and use them to inform our candidate design principles.

2.1 Sound Design Recommendations from Within HRI

While HRI researchers have so far not proposed any comprehensive, generalised guidelines on robot sound, some design considerations can be extracted from literature across the field. These considerations address design questions such as "What are the functions sound can fulfil in HRI?" "How could the different elements of a robot's sound be categorized?" and "How could different categories interact?"

For example, sound fulfils various functions in HRI scenarios. It is most commonly used to communicate affect [55], but can also help humans localise a robot in their environment [11], influence the perception of robot movement [42] or mask motor sound [51]. A broad categorisation of sound functions is proposed by Latuperissa and Bresin, who derive three general sound categories from an investigation into robot sound in film: inner workings, communication of movement, and expression of emotion [26]. The underlying design recommendation would then be to consider how deliberately and successfully a given robot sound design achieves all these above-mentioned effects, ensuring designers make full use of sound as a medium. A number of studies emphasise the need for designers to consider how unintentional motor sound affects HRI [20,31,48]. Sound intentionally emitted by the robot is not the only thing that communicates something to the listener. This encourages designers to take a more holistic perspective on robot sound, and be aware of all sound emitted by the robot and its environment to maintain control over what is being communicated. One common element across these recommendations is that they are often based on quantitative evaluations and do not feature perspectives of designers, who ultimately make the decisions on how a robot sounds.

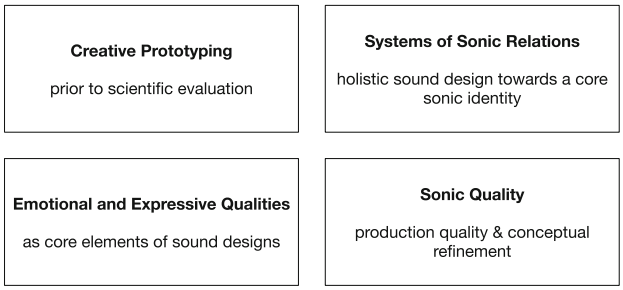

Fig. 2 A selection of key notions in product sound design, as reported by professional sound designers. Adapted from [23]

2.2 Product Sound Design–Robots as Objects to be Sonified

We can, however, find frameworks informed by the knowledge and experiences of practitioners in other fields. Viewing robots as objects to be imbued with sonic characteristics and behaviour allows us to draw connections to the area of product sound design. Work from that domain has been previously featured in HRI works, albeit with a focus on evaluation, rather than the design process [30].

2.2.1 Sound Design of Interactive Products

Hug and Misdariis [23] propose a conceptual framework for the sound design of interactive products, or, as they call them, *expressive artefacts*. They derive their framework from a combination of designerly and scientific sound design methods and evaluate it using expert interviews. A selection of key notions that emerged during their interviews with professional sound designers is shown in Fig. 2.

Creative prototyping is described as a fundamental aspect of product sound design, and meant to be the very first step towards designing sound for a product. This is a methodological concern, whereby designers prototype designs in an iterative free-form approach as opposed to making design decisions based on empirical findings, or psychoacoustic criteria. Scientific evaluation then provides a way to validate previously made design decisions, rather than dictating future ones.

Sonic quality is a core concern of sound designers, incorporating both production quality in technical terms and a design's rich and refined conceptual foundation. Interviewees stated that only this combination would result in high-quality designs with their own character, rather than sounds that merely reference established sounds of "heritage artefacts."

Emotional and expressive qualities should form a core element of every sound used, meaning that even for purely functional sounds a designer should ask, "What emotion should be expressed in this instance?"

Systems of sonic relations create connections between all sounds emitted by an object. A holistic sound design process therefore works towards a core sonic identity. As Hug and Misdariis put it, "sounds have to be part of a holistic design of an artifact and its process, in order to fulfill the potential role of sound to be 'the voice of things'" [23, p. 26].

While Hug and Misdariis' primary focus is on product sound design, the link to robotic agents is not far fetched, and their goal of providing actionable guidelines for sound designers aligns well with the aims of this paper. Applying their conceptual framework to robot sound design, one could propose the following design recommendations: (1) Robot sound should be informed by a combination of the sound designer's intuition and insights from behavioral studies. (2) The various elements of a robot's sound design should inform and reference each other to create a clear and recognisable core sound identity. (3) All robot sound, not just affective utterances, should aim to convey emotional and expressive qualities. (4) Robot sound should be based on a clearly defined conceptual foundation. It should be noted that while recommendations 1, 2, and 4 can be viewed as generally applicable regardless of application context, recommendation 3 applies to robots which, just like products, aim to delight their users with engaging and pleasant interactions. In applications where communication is purely functional, or where an attachment to the robot may even be undesirable, such as in military or rescue operations, designers may deliberately avoid emotional and expressive qualities. Carpenter, for example, notes how the deployment of humanoid robots in bomb disposal units could create attachments that "complicate emotional and ethical issues in terms of how human team members view the robot" [10, p 609].

2.2.2 Narrative Metatopics

Earlier work by Hug [22] presents a range of *narrative metatopics*, defined as "abstracted themes and attributes associated with narratively significant artifacts" [23, p. 28] (see Table 1). These metatopics emerged in structured group discussions around thirty films and games, where the sound component played a significant interpretive role. They represent a range of qualities inherent to an object or to an interaction with an object, which can be communicated through sound.

While Hug applies these metatopics in the context of product sound design, they too are applicable in the context of HRI. Notions such as *nature and judgement of artefact*, *manifestation of life*, and *qualities of control*, for example, resonate with HRI concepts such as robot perception, animacy, and agency. In practice, the metatopics can be used during the designer's concept creation process by providing metaphors that inform and guide the design process. This could, for example, be done by asking "What is the

Table 1 Narrative Metatopics proposed by Hug. Adapted from [22]

Nature and judgement of artefact	Structural states
Qualities of use	Manifestation of life
Qualities of control	Gesturality
Power/energy and its qualities	Transformation processes
Energy/power life cycles and dramaturgy	Temporal structure
Atmosphere, mood	

atmosphere/*gesturality*/*structural state* of this robot and how should sound communicate these qualities?"

2.3 Film Sound–HRI as a Narrative to be Scored

The interaction between a human and a robot can also be viewed as a narrative that is scored using sound and music. This perspective brings it in line with sound design practices in film and interactive media and thereby makes it possible to draw from conceptual work in those domains.

2.3.1 A Spectrum of Sound

A high-level perspective on sound in film can be found in esteemed sound designer Walter Murch's conceptual spectrum of sound [34], shown in Fig. 3. Murch places film sound across a one-dimensional spectrum ranging from language (encoded) to music (embodied) with natural sound design in between. Hybrid forms then comprise *linguistic* sound effects, such as a knock on the door, and *musical* sound effects, like musically embellished nature soundscapes. He defines *encoded* sound as sound, whose *meaning has to be extracted*, as opposed to *embodied* sound, which is *experienced directly*. A more thorough review of the terms encoded and embodied beyond Murch's application is beyond the scope of this paper. Murch suggests that five evenly spaced items across this spectrum result in the most transparent and at the same time information-rich soundscapes. It results in sonic environments that are "simultaneously dense and clear" [34, p. 20]. He also notes how a sound's position on this spectrum has implications for the spatial distribution it should have. An example he gives is the fact that encoded sound like dialogue or footsteps is traditionally centered, regardless of the location of the actors on screen, while more embodied sounds such as atmospheres or music are often freely distributed across the space to make full use of the spatial sound capabilities of the cinema environment. Essentially, as sounds move to the right of this spectrum, sound localisation becomes a more expressive sound design dimension, while sound on the left side is more restrained by listener expectations.

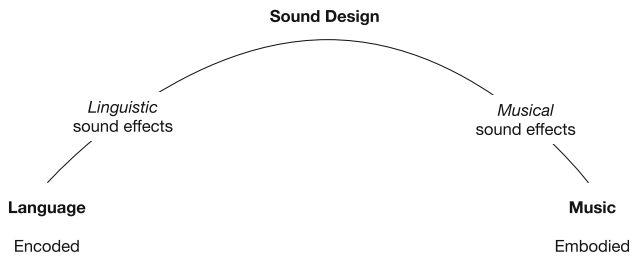

Fig. 3 Murch's conceptual spectrum of sound. Adapted from [34]

While this high-level perspective on film sound does not provide accurate categorisations for all types of sound, it is a useful tool to position sounds in context with each other. In the context of HRI, it provides a framework for looking not at individual robot utterances, but at soundscapes that accompany human-robot interaction narratives. For example, when looking at a robot utterance during an interaction, one might ask, which other four sound elements across the language-music spectrum could be used to fully score this scene. The spatial considerations are notable as well, as this framework provides an explanation for the dominance of single-source sound systems in the speech-oriented field of social robotics. It should be noted that the idea of *scoring* human-robot interactions with additional environmental factors may be enriching in some applications, but unnecessary and distracting in others. These two cases are illustrated in the case studies in this paper, where robot Cozmo (Sect. 3.1) features interactions that are fully scored with music emitted by a smart phone companion app, while the sound of robot Kuri (Sect. 3.3) is deliberately constrained to a small number of utterances to blend into the background.

2.3.2 Functions of Sound in Film and Interactive Media

Looking at sound in film and interactive media, we additionally find a wealth of functions sound can fulfil. Theorists in that domain have produced various collections of sound functions with various levels of detail [14,15,27]. The perhaps most thorough and methodical classification is proposed by Wingstedt, who assigns functions to categories and subsequently assigns these categories to classes [53]. For this paper, we will focus on four key classes, emotive, informative, descriptive, and guiding (see Fig. 4). To illustrate these four categories, it is helpful to imagine an example scene from a film: two people having a conversation in a room.

The emotive class uses sound to communicate and influence emotions. It contains the use of sound to describe emotions, describe relationships, and induce moods, among others. The example scenario could use music to underscore the conversation between the two people, conveying how they currently feel, how they feel about each other, what the emotional subtext of the conversation is, and how this conversation might impact their future behaviour.

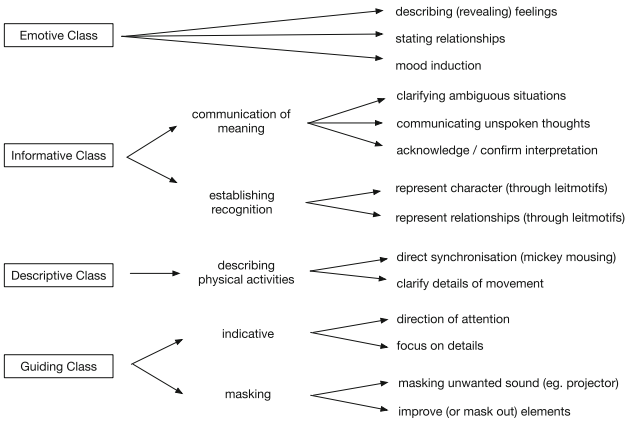

Fig. 4 A selection of Wingstedt's functions of sound. Adapted from [53]

The informative class uses sound to communicate additional information and context to the listener. It does so by communicating meaning and values, and by establishing recognition. In the example, this could be the use of historical soundscapes from outside the room that invoke a specific time period, or the use of metaphors like police sirens to indicate the neighbourhood the conversation takes place in and, by extension, the social status of the people involved. It could also involve the use of musical motifs for the individual characters, or musical material that conveys the wider significance of the conversation, and whether it has positive or negative connotations with regards to future events.

The descriptive class uses sound to describe settings and activities. It emphasises information that is already communicated through other channels and adds and clarifies information that would otherwise be ambiguous. In the example case, any events inside and outside the room could be sonified, describing environment, time of day, and season. It could also be used to clarify the actions of the two people, such as indicating one person's tight grip around a leather chair through creaking to indicate the force used and, by extension, how tense the person is.

The guiding class uses sound to guide listeners through the experience. It does so by emphasising elements which are important and de-emphasizing or hiding elements which are not. In the example scene, the breathing of one person could be amplified to focus attention on their inner state. In moments of silence, a subtle but noticeable room tone might be added to avoid complete silence, which, in a cinema, would draw attention to the sounds present in the viewing room itself, such as whispering in the audience or projector humming.

Analysing the various sounds emitted by robots through the lens of these functions provides several benefits. (1) It provides a pragmatic, goal-oriented view on a design process, helping practitioners more closely examine the motivation behind their creative decisions. (2) It ensures designers make full use of the communicative potential of sound, thereby avoiding missed opportunities. For example, a designer might ask what function is fulfilled by the sound they are currently integrating, or, in reverse, which roles sound could play in the scenario they are currently designing for.

3 Case Studies–Commercial Social Robots

We have so far covered a range of sound design considerations both from within and beyond HRI. This section examines the design process of three sound designers who created sound for commercial social robots (see Fig. 5). By conducting a detailed examination of the designers' process we aim to (1) gain insights on current practices in sound design for robots in real-world applications, (2) hear about the lessons learnt and challenges faced along the way, and (3) identify ways in which industry design work and HRI research could inform and enrich each other.

Our selection criteria for the case studies were (1) having a dedicated audio professional responsible for the robots' sound and (2) having budgets large enough to allow designers to develop concepts and iterate content over an extended period of time. We conducted qualitative interviews with the sound designers, asking them about their conceptual approach to the designs, and the creative and technical aspects of the design process. The interviews were subsequently analysed using thematic analysis [5]. However, it should be noted that thematic analysis of expert interviews commonly involves a higher number of interviewees which was not feasible in our case due to the small pool of suitable participants. Quotations in this section are taken from the interviews with the respective expert. Full transcripts are available as supplementary data.

Fig. 5 The robots discussed in the interviews include Vector (left), Jibo (middle), and Kuri (right). The sizes depicted are not to scale

3.1 Cozmo and Vector

Cozmo and Vector are two palm-sized robots initially created by manufacturer Anki, which is no longer operating. The robots are currently sold and maintained by consumer robot company Digital Dream Labs. Cozmo is a toy robot aimed at children and operated with a smart phone. It can recognize faces and play games with a user, among other things [38]. Vector is based on the same robot body as Cozmo, but features an integrated chip that allows it to run without an accompanying mobile app. It is targeted at young adults and older. Neither robots use speech and instead communicate through semantic-free utterances. The sound of both robots was designed by, and under the supervision of, Anki audio lead Ben Gabaldon. The following themes emerged from a one-hour interview with him.

Creating a believable **fictional character** was the core concern and main challenge throughout the design process for both Cozmo and Vector. For Cozmo, a common creative foundation among designers that guided all creative choices was not established before content generation. According to Gabaldon, it is "impossible to create compelling content" unless "you understand what the robot is." A key question that needs to be answered at the very beginning is therefore: "What is the fiction?" Learning from this during the design of Cozmo, Vector later had a character director, whose task was to answer this exact question.

Character and fiction were established through the use of metaphors. Cozmo was initially meant to sound like a "fine Swiss watch," in order to "invent" what is "under the shell," enhancing how users perceive the product quality of the primarily plastic robot. These ideas can also be found in the discussions around product and consequential sound [31,48]. The actual final metaphor used for designing Cozmo was a "bratty child" or "toddler," which then moved the sound design focus from "high-tech" to more natural sounding gibberish speech. Cozmo's sound is primarily comprised of processed utterances by a voice actor. This presents an interesting tension between communicating product quality and personality. In the case of Cozmo, personality was ultimately prioritised. In Vector, on the other hand, Gabaldon aimed for a sound design that can communicate both. Vector needed to use sound to differentiate itself from the similar-looking Cozmo. That, combined with the older target audience, required a different metaphor. Vector should be a "small foreign creature" and behave like fennec foxes or parrots, "intelligent creatures that can communicate in their own way." Vector was conceptualised as having a vocal tract made out of "moving small motors," allowing it to sound modern and foreign, while being in line with the character fiction. Vector's sound is largely built with processed recordings of a shoe shine machine.

Speaking of metaphors, Gabaldon also mentions a third, never released, human-sized robot by Anki. Due to its size and its target audience, adults, it was meant to be "more like a golden retriever" which would be a "larger, more intelligent creature that, instead of being startled by objects, would acknowledge them." The sound design approach was then planned to be more "smooth and selective."

Regarding the **design process**, Gabaldon notes how the robots' face and body animations, and sound were being developed at the same time. Audio production therefore needed to be fast and flexible, so that content could quickly be integrated into new animations. This enabled animators to present their work in context, which was necessary for their separate feedback processes. For Cozmo, this flexibility and speed was achieved by not creating isolated sound, but a large, modular set of sound events, "basic building blocks," which could then quickly be combined to create a broad variety of longer expressions. In Gabaldon's words, once a new sound was created, it was "accessible to any other animation so you organically just develop a larger and larger personality palette."

Sound design was initially done to videos of the robot, but the eventual mapping of sound events to movement was implemented using Wwise [2], an audio middleware used to implement sound assets into video games. This essentially separated the design process into two phases, a content production phase where individual sound assets were designed, and an implementation phase, where these sound assets where assembled into more high-level expressions and synced to robot behaviour. Gabaldon notes how he would occasionally return to the first stage to further develop or replace individual building blocks. Speaking about the evaluation of the designs, Gabaldon describes it as a "group effort." Creative choices were made in an iterative design process and evaluated by a small group of designers and stakeholders.

Among others, this iterative feedback process had to address one specific challenge. Gabaldon notes the key requirement of **emotive clarity**: making Cozmo and Vector communicate their personality through clear affective utterances. He mentions two challenges. (1) Affective utterances which were clear to Gabaldon were not clear to other members of the team. (2) Affective utterances that were clear when looking at the robot were not clear when the robot was out of sight. Previous work in HRI has shown how changes in context influence interpretations of affective utterances, see e.g. [40].

The team explored various solutions. An early idea for Cozmo was to take inspiration from non-verbal fictional characters like Star Wars' Chewbacca and Han Solo, and have a third party translate what Cozmo says. The robot could thereby make complex and unintelligible sounds, whose meaning would then be translated by an English-speaking "peripheral character." This would essentially remove the

need for emotive clarity, and allow Cozmo to sound as "unique and foreign" as desired. However, this approach was not pursued further, due to the resulting need to not only communicate Cozmo's fiction, but to "sell this whole universe" including additional supportive characters.

Instead, the process for achieving emotional clarity with Cozmo was as follows. Utterances were based on recordings with voice actors under the direction of Gabaldon, who notes the importance of directing actors to get "the right performance." Parallels can be drawn here to works in HRI, which aim to use the experience and intuition of creatives to inform utterance design. Savery et al. generate emotional prosody with a machine learning model trained on improvisations by vocalists [44]. Robot movement sonifications have been informed by actors [37] and musicians [16]. Gabaldon notes how this directing of voice actors is a skill that needs to be developed in order to get high-quality sound material to further process. Finished utterances would then be evaluated by other members of the team, both with the animations and isolated without context. The emotive clarity of the utterances was qualitatively assessed by small groups of designers, and refined over several iterations.

In summary, the design of the utterances was based on the experience and instinct of the sound designer, as well as on qualitative feedback from small groups of team members. HRI literature on the topic was not consulted, nor was quantitative evaluation performed.

During the development of Vector, Gabaldon explored various ways to have the robot's behaviour change over time and to introduce variation in the interactions that would result in "slightly different" robots for each user. He calls this **adaptive audio**. One idea was for the robot's utterances to become more complex and refined, as its perceived understanding of the objects in its environment developed over time. Gabaldon gives an example of the robot reacting to being too close to the corner of the table. After five occurrences of this it would identify the corner with a "rudimentary word," which then gets more refined after 15 occurrences. The robot "can start creating words around these [objects or events] depending on the number of times he experiences them." At this point in development, Vector's way of communicating was decided to be much less human-like than Cozmo, and this notion of language development was abandoned. Gabaldon believes, however, that with more development time the team could have found alternative ways to represent a noticeable and convincing development of Vector's language over time, even if that language was far removed from human speech.

Another application of adaptive audio was a more immediate variation of the robot sound. Vector has an internal parameter called "Stimulation," which represents its level of engagement with the environment. Calling Vector by its name sets the value to 1. Not interacting with the robot eventually lowers the value down to 0. This global parameter

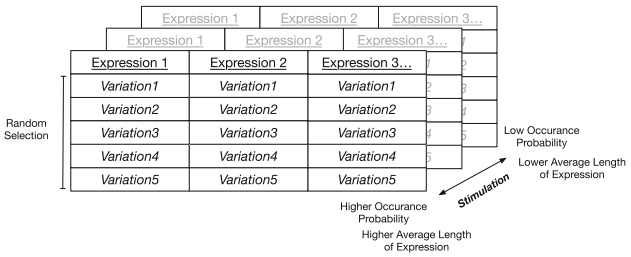

Fig. 6 The meta parameter stimulation affects the playback of expressions, changing probability of occurrence, and average length of expression, among others

affects, which of the many variations of Vector's various utterances are played back in any given interaction. As a result, direct interaction with Vector involves complex and attention-grabbing sound, while having Vector run in the background, would make him use shorter, softer, unobtrusive sounds that do not draw attention. According to Gabaldon, "people [...] might enjoy the engagement but as soon as they want to [...] work or do anything else it needs to get quiet." This mapping is shown in Fig. 6.

Vector was planned to have several of these high-level internal parameters which could be used to influence the audio on a global level. For example, the amount of an emotion like fear could then be used to apply a tremolo effect on all audio output, to "make it seem like he's shivering." Precisely defining Vector's character was ultimately prioritised over more expansive adaptive audio, and this was therefore not implemented. However, some of the ideas were expanded on in a patent filing [54].

Gabaldon differentiates between the above-mentioned adaptive audio techniques and **procedural audio**. The latter describes the real-time translation of internal robot data into sound. In the case of Vector, Gabaldon took data from the robot's head, lift, wheels, and screen to generate audio output. The motivation for this was to present the user with "interesting reactive sound," a complex audio system for a complex piece of technology. However, emotive clarity turned out to be a challenge, and much of the emotion and character meant to come across was lost. In Gabaldon's words, "[a scared robot's] fast-moving motors informing an audio system don't necessarily sound scared." Here, parallels can be drawn to findings by Frid et al., who explored what affective meaning participants perceived in the natural motor sound accompanying robot gestures [20]. More generally, work in HRI has explored real-time sonification of robot motion to affect vocal timbre [36], and utterance volume and prosody [4,45].

Vector's procedural audio system was eventually almost completely disabled, despite being fully implemented and functional. Gabaldon argues that "having an audio system that's responding to speed, mood, temperature, time of day

[...] doesn't really create interesting interactions." Instead, communicating a "compelling character" through "bespoke sounds" does. He does, however, note that access to more parameters of the robot's internal state, combined with more development time might have resulted in a procedural audio system that could compete with hand-authored content.

Cozmo and Vector's utterances can be separated into various **categories of sound**. However, for Gabaldon, all these categories constitute the "voice" of the robot and their core purpose is to create a believable character and communicate personality. Gabaldon divides Cozmo's sound into three categories. *Vocal utterances* are its way of speaking, sounds associated to movement on Cozmo's *screen* communicate facial expression, and sounds accompanying the robot's *lift* movement communicate body language. Vector has four categories of sound: *head*, *lift*, *treads*, and what Gabaldon calls *the emotional layer*. While the former three are again closely linked to the robots screen and body movement, the latter is not synced to any movement, but rather an "a lot softer" layer used to emphasise the robot's emotion during rare, "impactful moments with people." For example, if Vector recognises a user's hand, it crawls up on it, and lets out a "synthesized, tremolated pattern," similar to a pet sighing. All the above-mentioned categories still each have their own sets of variations, as illustrated earlier in Fig. 6. This means they get softer, shorter, and less dense when the robot is not directly being interacted with.

Generally, every action of the robot comes with a corresponding sound, or, every sound the robot makes is synced to facial or body movement. The sole exception of this is Vector's *emotional layer*. According to Gabaldon, every time the robot does something, sound is used to "fill out that experience," giving the sound designer "subtle opportunities to communicate" character and personality. When used in subtle ways that do not get "obnoxious over time", sound can then be used to make the robot's fiction believable by creating an "enhanced reality."

When asked about the relationship between robot sound an **sound in the environment and background**, Gabaldon notes that the possibility of broadcasting robot sound to IoT-enabled sound sources like smart speakers in Vector's environment was considered during development. However, the team came to the conclusion that emitting robot utterances from anywhere except the robot itself "breaks the fiction." A believable character requires convincing and consistent sound localisation. In the case of Cozmo, there is a fully interactive musical score that accompanies the mostly game-based interactions, which is played back via the robot's accompanying mobile app. However, Gabaldon notes that an "understood and parallel" music system is different from having the robot communicate through different sound sources. In his opinion, a convincing character would "respond to" external sound events like music, rather than "sharing" that music as its own content. The robot should be "responding to the world with you and not creating it for you."

This notion of *responding* to external stimuli *with* the user even if those stimuli are technically also triggered by the robot was pursued further. For example, the team looked into beat detection to sync robot movement to external music (note the parallels to work by Hoffman and Vanunu [21]). Another illustrative example is how Vector reacts to more functional questions like how the weather is outside. Asking it for the weather makes weather information appear on the robot's screen in a manner that pushes vectors face into the background. In Gabaldon's words, "he doesn't show you the weather with his own abilities. The weather shows up and he responds to it. He's still vector inside of this [...] external utility."

3.2 Jibo

Jibo is a foot-tall social robot for the home created by Jibo, Inc. under Cynthia Breazeal. Like Anki, the company is no longer operating, and Jibo is currently not available to the general public, but to businesses and institutions in healthcare and education. The robot communicates by using language and additional non-verbal sounds. The sound of Jibo was designed by, and under the supervision of, audio designer Jeshua Whitaker. The following themes emerged from his written answers to a series of questions.

Jibo's sound can be categorised into three main pillars: text-to-speech (TTS), semantic-free utterances, described as "jiboease," and user interface (UI) sounds. The dominant role of the TTS system was to convey information. Utterances would precede or follow Jibo's sentences in order to communicate character and emotion. Jiboease could also appear outside of user conversations in the form of "idle chatter," which Jibo would emit when left on his own. Finally, UI sounds would sonify user actions on Jibo's screen such as moving sliders, pressing buttons, and sonify system events like startup and shutdown. According to Whitaker, UI sound design provided a "subtle way to communicate with users when interfacing with Jibo via his menus."

The core function of Jibo's sound was "to **create a character** that users could connect with." The robot's personality was to be "joyful, smart with a hint of sarcasm, and some dry humor." The robot's speech engine allowed for some control over prosody, patterns of stress and intonation in the language, but this proved insufficient to convey an expressive personality and the bulk of character communication was therefore done by the semantic-free utterances. These were specifically developed with that function in mind, allowing Jibo to go beyond the limited prosody of the TTS engine to "communicate more colorful emotions such as sadness or excitement."

Communicating emotions clearly and unambiguously was a recurring challenge throughout the design process, both (1) when utterances were heard on their own, and (2) when utterances were combined with other modalities, such as graphical screen elements or gestures. User surveys during development showed that semantic-free utterances would reliably convey the desired emotion around 40% of the time. As part of a multi-modal expression system, they were more successful, reaching a 70% success rate. This, however, was additionally helped by restricting the conveyed emotions to simple, broad categories, such as happiness and sadness, rather than more nuanced notions like sarcasm. This resonates with findings by Read and Belpaeme, who demonstrated how humans tend to interpret robot utterances categorically, drawn towards prototypical emotions [41].

Jibo's **sound design process** can broadly be separated into a creation, and an implementation phase. The process started with early concept work which was informed by popular designs in science fiction, like R2D2 and Wall-E, and also included conversations with these two characters' sound designer, Ben Burtt. The first six months of the design process did not involve the physical robot. Instead, sound choice and timing were sketched out using linear video clips ("to picture"). Designs were iteratively validated using user surveys, which, according to Whitaker, "helped simplify how we use the robot speech sounds and gave more focus to what worked and [what did not]." Finally, sounds would be integrated into the physical robot, adjusted to the acoustical properties of Jibo's sound system, and mapped to various behavioural parameters. Whitaker notes, that in an ideal scenario, more time would have been spent on conceptual work at the very beginning of this design process to more precisely "outline the use case of the robot and see where audio and communication slots in."

Whitaker faced various **challenges** throughout the design process. His sound design had to take into account limitations of Jibo's sound system, two speakers on the left and right of the robot's head. Besides frequency limitations - frequencies below 170 Hz had to be avoided - he notes that users would not receive direct sound from Jibo's speakers, as the speakers would rarely face the listener directly. Instead, users would hear the sounds' reflections from the walls around the robot. This would "throw some users off since we tend to locate objects based on where our ears perceive the first sound." His suggestion to address this problem is to use a forward-facing speaker for speech output. (See also Brock and Martinson's work on "auditory perspective taking" [8].) Jibo's head additionally acted as a resonating chamber, meaning sounds had to be processed to account for this before implementation. Another challenge was the previously mentioned ambiguity of the robot's semantic-free utterances. Whitaker notes that if he were to design sound for a new social robot he would rely on recent advances in text-to-speech engine capabilities to shift the task of communication emotion exclusively to speech, removing the need for a semantic-free robot language altogether.

Implementing sound into Jibo and making it **responsive to user actions** was done in several ways. All non-speech sound is stored as sound files, meaning no sound is generated in real time during interaction with the user. Sounds are tagged with meta-data, allowing narrative designers to pick appropriate sounds for various situations. Each type of sound consists of a group of subtle variations, from which one is randomly selected during playback in order to create variety. Jibo also has an internal parameter which keeps track of how recently somebody interacted with him. The earlier-mentioned "idle chatter," utterances outside of user conversations, are linked to this internal parameter, affecting Jibo's sound when he is ignored for longer periods of time. Whitaker notes, that with today's technology, he would have aimed for a "smarter emotional system" which could have taken in real-time parameters like environment sound, light, or a user's facial expressions to inform the robot's sound output in a more responsive way.

3.3 Kuri

Kuri is a social robot for the home. It is half a meter tall and was designed and produced by Mayfield Robotics. The company is no longer operating and the robot never made it to market. Kuri is said to have an expressive personality which it communicates solely via semantic-free utterances. The robot's sound was developed and realised by Connor Moore, audio user experience lead for California-based CMoore Sound. The following themes emerged from a 1-hour interview with him.

Moore notes three core **design goals** that guided the development process. (1) Kuri should express himself in a positive and welcoming way to gently integrate into people's homes, despite being what is essentially a *foreign object*. (2) Kuri should be able to effectively communicate basic concepts like positive and negative statements, and simple emotions like happy or sad, and have those quickly and intuitively understood by the humans around it. (3) The sound of Kuri should reflect the Mayfield Robotics brand, providing a through line across potential future robots released by the company.

Kuri was meant to communicate a delightful **personality** that was "bright" and "cheerful," being "a joy to have around in your home." Its sound should communicate "organic" qualities with a certain degree of "imperfection," while also not being too human, as designers were concerned that too high a degree of anthropomorphism would make the robot appear "smarter than it should," running the risk of over-promising and under-delivering functionality.

Moore took all the above-mentioned goals and considerations into account when designing **Kuri's voice**, and eventually

arrived at the following design. The sound of Kuri's voice is modeled after an mbira, an African thumb piano. The sonic characteristics of the instrument were digitally reproduced in a process called re-synthesis, which allowed Moore to then apply arbitrary pitch and volume contours. He could then "model [...] human intonation" while preserving the "organic quality and [...] imperfection" of the source material. Additional processing of the sounds would include cyclical modulation of the sound volume to create purring or growling. Some expressions included additional sound effects like the beating of a heart, and alert sounds used for waking alarms. While these sounds were separate from the robot's central sound source, the modelled mbira, they were integrated in a way that made them part of the robot's language and kept them "based in the organic." The rationale for this was consistency in the robot's sound characteristics. Moore notes that having a "consistent thread" was essential for the Mayfield Robotics sound branding and was therefore a "primary goal" of this sound design process. Kuri's sound needed to be consistent and at the same time effectively fulfill its various functions. This same design approach can be found in the framework discussed in Sect. 2.2, particularly regarding notions of a *core sonic identity*. The timbre can also be seen as a leitmotif (see Sect. 2.3.2) that is present throughout all interactions with the robot.

Moore's **design process** consisted of four stages: (1) Discovery and Strategy, (2) Creation, (3) Iteration, and (4) Optimization.

Phase one involved research into the various ways robots communicate, both in current real-world implementations and in fictional depictions. References included robots Vector and Jibo, and historical and more current popular fictional robots from science fiction. This phase also included research into the possible ways the Mayfield brand could be communicated.

Phase two involved extensive prototyping of a broad range of possible sound design approaches. This enabled Moore to explore and quickly assess different directions for the robot's sound, particularly in regard to the previously established design goals. Moore explored, among others, various acoustical sources like wind instruments, processed and layered vocal recordings, synthetic UI sound based on simple sine tones, and musical phrases that used melody and harmony to communicate, as opposed to sounds resembling human prosody. These diverse approaches could then be compared to choose a general direction for the sound design of the robot. After synthetic approaches "didn't seem quite rich enough," the team arrived at the above-mentioned mbira as rich and flexible source material.

Phase three was the iterative extension and refinement of this approach, involving continuous content production by Moore, feedback from Mayfield executives, as well as user testing with groups sourced both from within the company as well as externally. The key question in this endeavour was whether Kuri's sound could "maintain that consistency and be functional at the same time," whether users would perceive a common thread throughout the robot's various utterances, and whether these utterances could effectively communicate their intended content.

During the Optimisation phase, the final set of sound assets was played back through the physical robot and adjusted to take into account the acoustical characteristics of Kuri's loudspeakers. Moore notes, however, that it was crucial to already involve the physical robot early in the design process, as this allowed him to (1) consider loudspeaker characteristics during early sound design ideation, and (2) assess how different sound approaches would play out in context. Finally, Mayfield's engineers would integrate the finalised audio assets provided by Moore.

Moore's background in sound branding and audio user experience design is reflected in his **design approach**, particularly regarding notions of **expression, immersion, and variation**. Unlike the two other sound designers with game-audio backgrounds, Moore was not involved in the final implementation of sound assets and Kuri does not feature the sonic variation commonly found in video game sound to create a believable fiction, or an immersive experience.

Instead, Moore favors a "simple design language" in line with what he calls "considerate or polite design." To him, deep variation can be useful in avoiding annoyance resulting from excessive repetition, but also runs the risk of becoming confusing. He notes that this choice needs to be highly considered and "very thoughtful in its execution." As an alternative, a language that is deliberately constrained and simple is a powerful way to introduce sound cues that are clear and intuitive, as well as "respectful to our environments;" sounds that can "exist in our spaces without taking up too much space." When asked how designing for Kuri differed from more conventional product sound design, he emphasizes that the robot's language is significantly more "playful" than products like mobile apps. In his opinion, product sounds should be "simple," "reserved," and "sophisticated." In contrast to that, "Kuri was much more playful than a lot of products should be."

When asked about **constraints** he faced during the design process, Moore mentions that a step-wise integration of variation parameters was to be considered at a later point in Kuri's life cycle, but this was ultimately halted when Mayfield ceased operations. Moore notes that the role of variation should have ideally been thoroughly and comprehensively mapped out during the early conceptual phase. He also notes that he would have welcomed having additional time during the early design phases to explore more directions. Another constraint Moore faced was Kuri's loudspeaker capabilities. Using loudspeakers which could comfortably reach the frequency range between 200 Hz and 500 Hz would have pro-

vided Moore with a broader sound palette to work with. He notes that lower pitch ranges "tend to be warmer and more inviting and welcoming," as well as "more polite and more soothing." Kuri's technical specifications therefore impacted its ability to convey character, personality, and emotional expression.

Kuri's **affective communication** is limited to a small number of core emotions, like happiness, sadness, and, in rare occasions, frustration. These expressions, called "Romojis," are then further differentiated by intensity, meaning Kuri could, for example, express pure joy, or slight sadness. The utterances were then combined with the robot's expressive eyes and a light source in its chest to create multimodal affective expressions. Mayfield's user testing showed how this combination allowed Kuri to effectively convey the intended emotions (see also Löffler and colleagues' work on the multimodal interplay between color, motion and sound [28]). When asked whether equivalents to Kuri's affective utterances can be found in more conventional product sound design, Moore states that he sees considerable common ground between the two. In his words, "you're always trying to convey some type of emotion through a product or a brand sound." Conveying emotional qualities through all aspects of the sound design is also mentioned by Hug and Misdariis in their sound design framework discussed in Sect. 2.2.

4 Candidate Design Principles

This section presents the core contribution of this paper: nine candidate design principles for robot sound, shown in Fig. 7. The principles are assigned to the following themes:

- **Fiction**–the conceptual foundation of the robot sound design
- **Source**–the perceived cause of sound and the physical location of the sound source
- **Scope**–the elements making up a comprehensive sound design
- **Interactivity**–the way in which robot sound reacts to the environment and develops over time
- **Content Production**–the iterative process through which audio assets are created and evaluated

Throughout this section, we present and describe each principle, addressing (1) *what* it is, (2) *why* it is relevant, and (3) *how* it could be implemented. The latter two points make reference to the frameworks in Sect. 2, the case studies in Sect. 3, and related work in HRI. We offer these principles to guide designers in creating rich, refined, and comprehensive sound designs for robotic agents. We welcome these principles being challenged and do not see them as rules to be followed, but rather as a framework to assess new and existing designs, and help in identifying missed opportunities.

4.1 Fiction

Fiction is a key element in robot sound design, and comprises the need for a clearly defined character as well as working towards a believable physical object.

Designing for a clearly **defined character** means teams have a clear and shared idea of the character and personality they are creating sound for. This ensures all designed sound is in line with the core fiction of the character, allowing that character to be communicated more effectively. This design

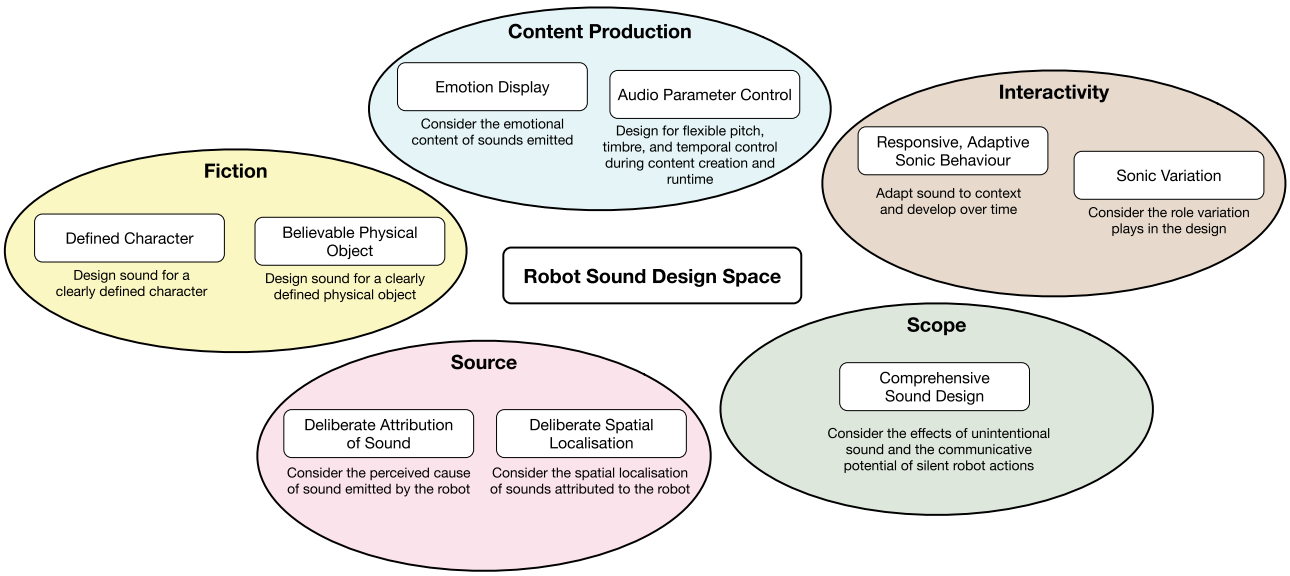

Fig. 7 Robot sound design space with candidate design principles

principle is a core aspect of many of the themes presented in this paper. While not aiming to create believable fictional characters, Hug and Misdariis' work in Sect. 2.2 emphasises the importance of having sound design draw from a conceptually refined foundation that can inform all design decisions throughout the production process. Whitaker mentions the value of a comprehensive conceptual foundation as the first step of the design process. In Moore's design, this foundation takes into account core characteristics of both the robot and the brand, and is communicated through a finely tuned sound set that consistently conveys a clearly defined identity. Gabaldon laments the initial lack of that foundation for his robot sound, which made it challenging to create a large, coherent set of sounds that effectively communicates one character and personality. Hug and Misdariis describe this as "Systems of Sonic Relations." Further, their Narrative Metatopics (Sect. 2.2.2) provide a broad range of possible notions that can inform the design concept. Applying the principle during the creation or evaluation of robot sound would, among others, involve the following questions. What is the core fiction of this robot character? What is a possible history? What are possible personality traits? How could this fiction be communicated through the robot's utterances, movement sonification, and other sound? What types of sound would contradict this character?

Teams should also have a shared idea of the robot as a **believable physical object**. This means getting a complete picture of the visible and hidden robot characteristics, like size, shape, and materials, and then considering how these characteristics are communicated through sound. It should be considered how sound either emphasises the characteristics, like giving a deep voice to a large robot, or how sound perhaps communicates something different, like sonifying the movement of a plastic robot with metallic sound. Communicating physical robot characteristics through sound is another common theme in the literature and case studies presented in this paper. Wingstedt notes how sound in film is used to describe and clarify physical settings and activities (see Sect. 2.3). Gabaldon speaks of "inventing" what is "under the shell" of a robot. Moore and Whitaker mention how different aspects of robot embodiment and sound inform each other's design and emphasise the need to iteratively asses sounds played through the robot itself. Various works in HRI support this notion, as well. Read emphasises that robot utterance design needs to take into account robot embodiment [39]. Moore comes to the same conclusion after investigating the perceived appropriateness of robot voices on different embodiments [32]. For the design of social robot Miro, Moore and Mitchinson take inspiration from mammal vocalisations to create a real-time vocal synthesiser that generates believable utterances in line with the appearance of the animal-like robot [33]. Much of the work on consequential sound is based on the idea of communicating robot characteristics through motor sound (e.g. [30,31,48]). Applying the principle during the creation or evaluation of robot sound would, among others, involve the following questions. What information, whether factual or made up, about the robot's materials and physical capabilities should be communicated? How could sound support this? What types of sound would contradict the physical attributes we are trying to convey?

4.2 Source

Sound sources are perceptually relevant elements of a robot's sound design, both in terms of the perceived causes of sound (Who is addressing me right now?) and in terms of the sound localisation (Where is the sound coming from?). In both of the following principles, sound plays a significant role in drawing a border between the robot and its environment. Designers should therefore consider the real and perceived sources in their robot sound and how these support, or distract from the sounds' intended message.

Deliberately attributing sound to specific sources means to consider the perceived causes of any sound designed by the robot. One might differentiate, for example, between sound "spoken" by the robot to address the user, UI sound emitted by the robot in response to some external input (e.g. receiving a message), and artificial movement sound emitted by the robot as part of its fiction. In robots Cozmo and Vector, these distinctions are drawn clearly. Robot utterances are different from an "understood and parallel" music system, and different from notification sounds associated with external stimuli. Jibo draws distinctions as well. While speech and robot utterances are part of the robot's language system, UI sounds are part of the robot's screen interface and adhere to different language conventions, namely those of appliances, smartphones, and tablets. Both Whitaker and Gabaldon note the importance of maintaining clear sound cause attributions. Jibo's loudspeaker placement sometimes places speech away from the robot's mouth, namely at walls around the robot, which is described as a design flaw. Cozmo and Vector both had the chance to emit sound from loudspeakers in their environment and did not, in order to maintain their fiction. Moore, on the other hand, makes the conscious design decision to derive all robot sound - utterances, UI, alerts - from a common source material, prioritising a consistent sound experience over differentiating these sub categories. Applying this principle during the creation or evaluation of robot sound would, among others, involve the following questions. Who or what is the perceived emitter of this sound (the character, third parties, environment)? Are these different causes of sound sufficiently distinguishable from each other, or are they meant to blend?

Deliberate spatial localisation is one of the key actions designers can take to attribute sounds to different sources. It involves making conscious decisions on exactly where the

robot emits sound, both on the robot itself and in the robot's environment. Possible speaker locations can, for example, include the robot's face and body, peripheral IoT devices with audio capabilities, and smartphones and wearables. Whitaker suggests using different speakers for different types of robot sound, by, for example, emitting UI sounds through Jibo's stereo speakers to the left and right of the face, while emitting speech from a central speaker close to the robot's mouth. In Cozmo, Gabaldon spatially separates robot utterances and musical score, placing the former exclusively with the robot, and music exclusively away from it. Parallels can be drawn to the Spectrum of Sound by Murch, shown in Section 2.3.1. He notes how listeners have clear expectations of the localisation of sound directly attributed to a source (like language or footsteps), whereas musical sound can be spatialised more freely without feeling unnatural. The perceptual relevance of sound localisation in HRI has been explored by Cha et al., who investigated how robot sound can help humans localise it in their environment [11]. Beyond that, the role of spatial sound as well as the attribution of sound to different sources is comparatively underexplored. More broadly, Wingstedt's Guiding Class, described in Sect. 2.3.2 suggests how sound can be used to direct attention to different objects or aspects of an interaction. With a parameter like sound localisation, these effects can be applied in the spatial context of HRI scenarios. Applying this principle during the creation or evaluation of robot sound would, among others, involve the following questions. Does the robot emit sound that is not meant to be attributed to it? Is the physical source of the sound in line with its perceived source? Can we communicate something through spatial localisation that might help with our message?

4.3 Scope

Comprehensive sound design considers all sound that *is*, and that *could be* emitted by the robot as part of the design process. Every robot action is an opportunity to communicate through sound and all robot sound communicates something. This means that the sound of a robot is not limited to the audio assets that were implemented into it, but instead it is an accumulation of all sound emitted by the robot, both deliberate and unintentional. Designers should therefore consider (1) what message any given robot sound conveys, and (2) what message it should convey about robot character, personality, state, and others. Comprehensive designs use a wide variety of sound or deliberate silence to support, augment, and enhance all robot actions, such as movement, facial animations and gestures. This design principle is derived from both industry practice and findings in HRI. For example, Frid et al. find a disconnect between inherent robot motor sound and the emotional content of expressive gesture [20]. Frid and Bresin therefore later blend artificial sound with robot motion sound to improve the clarity of said gestures [19]. Other unusual applications of sound include synchronising robotic motion with music extracts [4] and accompanying robot gestures with infrasound [49]. Whitaker uses sound to give a voice to Jibo during voice interactions, and to provide subtle context and guidance during touch interactions with the robot's screen. In Gabaldon's work, every single robot movement and gesture is coupled with sound. Even when used subtly, sound reinforces core messages about robot character and personality, and the rest of the robot fiction. The scope of sound design in the three case studies is shown in Fig. 8. Applying the principle during the creation or evaluation of robot sound would, among others, involve the following questions. Are there aspects of the fiction that are currently not communicated through sound? Are there robot actions that are currently not enhanced through sound? What do the robot actions not deliberately accompanied with audio sound like? Are these sounds in line with what should be communicated? How could sound distract from characteristics we do not want to emphasise?

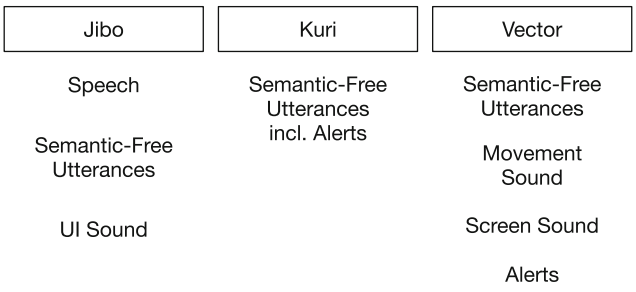

Fig. 8 Scope of sound design of three robots from the case studies

4.4 Interactivity

Interactive sound can take various forms in human-robot interaction. Sound can adapt to individual interaction scenarios and change and develop over extended periods of time. It can also feature subtle or more obvious variation in response to both internal and external parameters. In the two following principles, sound allows robots to communicate animacy and awareness of their environment, and adjust to various interaction contexts. Designers should therefore consider the role interactivity plays in their design and how this interactivity could be achieved.

Responsive and adaptive sonic behaviour allows robot sound to adapt to context and develop over time. By creating sonic behaviour that responds to the environment in both obvious and subtle ways, designers can address functional concerns such as speech intelligibility or listener fatigue, and, more generally, create richer interactions and more believable autonomous behaviour. Examples include continually adjusting parameters like volume and timbre based on environment

noise levels, or selecting sound assets from a pool of options, depending on parameters like "time since last interaction," or "number of humans present." This design principle can be derived from both existing work in HRI, as well as the case studies. In the perhaps most functional example of making robotic sound responsive to user behaviour, Brock and Martinson propose the notion of "auditory perspective taking" to optimise a robot's speech intelligibility. They suggest four measures adaptive systems might take: facing the listener, adjusting speaking volume, pausing when environmental sounds are too loud, and moving to another location when environmental sound persists [8]. Schwenk and Arras continuously sonified the distance between a human and the robot [45]. Both Whitaker and Gabaldon implement "idle chatter" for their robots and link its occurrence to how recently the robots were interacted with. As a result, the robots exhibit human-like behaviour by keeping themselves entertained when on their own. By additionally emitting these utterances in the form of unobtrusive peripheral sounds, the robot maintains an auditory presence while seamlessly moving in an out of focus according to user requirements. Gabaldon also suggests adaptive behaviour with a longer time frame. By keeping track of past experiences and slowly developing specific utterances around those, the robot essentially forms a language and thereby emphasises how it keeps track of past experiences and how it learns from them. Applying this design principle could involve the following questions. How does sound need to be adjusted during interaction scenarios to create seamless interactions? How could interactive sound be used to emphasise a robot's awareness of its environment? How could adaptive sound be used to emphasise that the robot becomes familiar with things and people over time?

Creating **variation** in sound by drawing from internal or external information provides a way to continuously modify sound material to avoid repetition and make robot sound more closely resemble real-world sound sources. Compared to the above-mentioned adaptive and responsive sound, variation works with a more low-level access to the audio material. One example is Vector's motor sound sonification, which is influenced by movement speed and direction. Another example would be Whitaker providing sets of utterances with subtle differences between them, which then randomly get selected during interaction scenarios. Both Gabaldon and Whitaker utilise their backgrounds in game audio and use variation to create a more believable character. If we suspend our disbelief and view a robot as a living thing with a personality, we may not want to think about the loudspeakers and sound assets involved in the communication process. Instead, we may prefer a fiction such as, for example, "the robot has a small, metallic vocal tract and cannot stop giggling." Variation allows fictions like that to be conveyed more convincingly. An alternative, product sound-oriented approach is demonstrated in Moore's work on Kuri. By deliberately constraining variation and focusing on a smaller set of sounds that are designed to be unobtrusive and timeless, the robot's language avoids the risk of becoming unnecessary complex. Considering the role of sonic variation during the creation or evaluation of robot sound would, among others, involve the following questions. Does the robot sound benefit from imitating the variation present in real-world sounds, such as speech, acoustical instruments, or noises? What parameters, such as data on robot behavior or the environment, are available to be used to introduce sound variations?

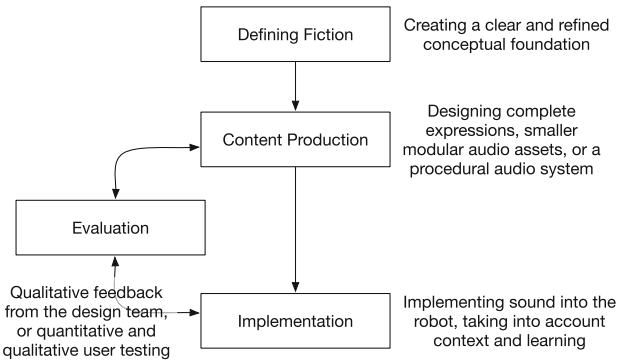

Fig. 9 Overview of the sound design process, comprising concept creation, content production, iterative evaluation, and implementation phases

4.5 Content Production

Content production relates to the way audio assets are created and evaluated. The most relevant design considerations are the design process, the display of emotions, and the control of audio parameters. An archetypal design process derived from across the case studies and design frameworks presented in this paper is shown in Fig. 9.

Emotion display is the dominant application of semantic-free sound in HRI research (see [55]). From a sound designer's perspective, sound can directly communicate the emotions of the speaker, i.e. the robot, but it can also more generally communicate emotional qualities through music or sound design with musical elements, as common in film and product sound. Designers should therefore consider the emotional content of sounds emitted and what effect these should have on the listener.

Across the case studies, the challenge of designing clear and robust emotion display is tackled in various ways. Both Whitaker and Moore use a relatively small palette of emotions and create and quantitatively evaluate their designs over multiple iterations. Moore additionally aims to convey emotional qualities through all sounds emitted by the robot, not just affective utterances towards the user. Gabaldon puts considerable effort into using semantic-free utterances to clearly communicate a rich palette of moods and emotions felt by the

robots Cozmo and Vector. He goes through a large number of iterations and collects qualitative feedback from the design team, as opposed to the general public. He also creates a procedural audio system, but later deactivates it because it does not meet the design goals of clarity and depth. In these cases, the path towards successful emotion display through sound therefore seems to be meticulous hand-authoring, thorough, iterative evaluation, and restraint in the range of emotion displayed.

Despite this, reliable affective communication using semantic-free utterances remains a challenge across the case studies and in the HRI research community. Next to the various methods described above, some HRI studies use machine-learning approaches to procedurally generate emotion display. One notable example of this is Read's work on using neural networks to learn mappings between an affect space and utterance parameters like frequency, speech rate, pause ratio, and rhythm, among others [39]. More recent and ongoing work by Savery and colleagues looks into how machine learning models can be trained to use musical improvisation to create what they call *emotional musical prosody* [43]. However, one could argue that these types of utterances can be a valuable tool to create richer and more engaging characters, even if the content of the communication is vague or unpredictable. Pelikan et al. speak of an "inherent vagueness of emotion displays" and how these are still valuable beyond the deliberate communication of specific emotions [38, p. 461]. When asked about ideal robot communication, Whitaker notes the growing accessibility of emotional speech synthesis and suggests making emotion display exclusively a component of verbal communication. He then suggests using semantic-free utterances for entertainment purposes only.

Considering the role of emotion display during the creation or evaluation of robot sound would, among others, involve the following questions. Is accurate emotion display a requirement for successful interactions, or can more ambiguous emotions be used simply to convey a richer, more engaging character? Will a more restricted set of archetypal emotions fulfil functional requirements of the interaction? Are there sufficient resources available to work towards refined and robust emotional communication through hand-authoring and extensive evaluation cycles. Are there successful procedural approaches? Which other modalities are available to support emotion display?

Audio parameter control relates to the various ways sound designers can shape sound characteristics both during production and after robot deployment. While audio parameter control seems like an obvious aspect for a sound design process, robot sound has specific requirements. For example, a designer might want to choose an certain timbre to convey general information about the robot (e.g. the mbira forming Kuri's core sound material), but then combine this with arbitrary pitch contours to create prosody. Having a robot's sound react to context and environment also requires control of at least some audio parameters. Designers should therefore consider how to design robot sound in a way that maximises pitch, timbre, and temporal control.

Audio parameter control can be approached in different ways, including (1) procedural generation, realised either through explicit mappings between sensor data and audio parameters, or with machine learning techniques, (2) hand-authoring of small modular sound assets that are then combined in an interactive audio engine like Wwise, or (3) hand-authoring of complete utterances, limiting real-time control to high-level controls like start-time and volume. Moore uses re-synthesis, synthetically recreating the timbre of an mbira, to get flexible control over the sound material. This allows him to communicate emotion and intent by emulating human prosody, while at the same time using specific sound characteristics as Kuri's sound identity. Gabaldon takes two approaches to control over the sound material. One is generating audio in real time based on internal robot data. Another is hand-authoring modular building blocks by hand to be able to later flexibly reassemble them for different interaction scenarios. He eventually chooses hand-authoring over a procedural approach due to the former's increased emotional clarity and better communication of intent. Procedural HRI approaches focusing on emotion display were mentioned in the previous section. Another notable procedural approach is found in Schwenk and Arras' real-time sonification of user distance [45]. An approach utilising flexible re-combination of small modular sound assets is found in work by Jee et al., who synchronize musical utterances with a robot's motion trajectories to make the emotion expression more effective. To this end, they design the musical structure of their utterances in a way that allowed for repeating short sections of the music. Utterance length can then be adjusted to coincide to various robot movements without running out or being cut off [24]. However, unlike in the case studies, these studies place little emphasis on using specific timbres that have a larger communicative significance.

Considering approaches to audio parameter control would, among others, involve the following questions. Which sound design approaches provide appropriate levels of control over the chosen timbres and pitch contours? What level of real-time control is required for the robot to have responsive, and adaptive sound, and what sound characteristics are involved? Is the chosen sound characteristic and degree of real-time control best realised using hand-authoring or procedural approaches?

5 Limitations and Future Work

When considering these candidate design principles, it should be noted that they are based on a limited number

of case studies, meaning they are based on the experience of a small group of designers creating sound for a small number of robots. While we believe that many of the design considerations in this paper are valid and applicable to any robot, insights into the design process for different robot embodiments and application areas may lead to the addition of new, or the adjustment of existing design principles. As noted in Sect. 3, the pool of sound designers who have deep experience in this domain is currently still small, and we hope to expand the pool of case studies in future work.

One particular challenge designers face is combining the sound characteristics dictated by a design's conceptual foundation with the interactive capabilities required by the HRI context. And, while meticulous hand-authoring in combination with extensive amounts of explicit sound mappings is a viable, albeit resource-intensive way to achieve this, machine-learning approaches may provide an alternative path. These types of procedural audio generation could allow designers to combine the timbres and sound qualities required for creating interesting sounding characters, with the interactive capabilities required for fluid and context-aware human-robot interactions.

6 Conclusion

The contributions of this paper are twofold: (1) we presented a detailed examination of the sound design process of three commercial robot case studies, and (2) we combined insights from these case studies with existing design frameworks beyond HRI, and findings from studies within HRI to propose candidate design principles for robot sound. Neither of these contributions have, to our knowledge, been made in prior work on robot sound. An examination of design frameworks in the areas of product sound design and film sound indicates that many of the concepts established there can be applied to robot sound to create richer, more conceptually refined designs. The themes emerging in conversations with the sound designers of Jibo, Cozmo and Vector, and Kuri reflect notions both from these HRI-external design frameworks and findings in HRI. We structured our candidate design principles along the themes *Fiction*, *Source*, *Scope*, *Interactivity*, and *Content Production*, providing a high-level overview of the challenges and opportunities of creating high-quality robot sound.

This provides sound designers with a broad range of recommendations to consider when creating sound for the HRI context, including (1) what elements of robot behaviour could be sonified, (2) what functions different types of sound may fulfil, (3) how different embodiments should be considered when designing appropriate robot sound, (4) how designs may be adjusted to work with constrained loudspeaker characteristics and placement, and (5) how sound assets may be created, which can respond to the interactivity inherent to HRI scenarios. Ultimately, sound designers create successful human-robot interactions by finding elegant solutions for the challenges imposed by their robot's specific application context, and by working within the various constraints they face when embedded in an interdisciplinary team of engineers, product managers, and other stakeholders. We hope our work supports them in this process, acting as a starting point for the sound design of future robots and a means to analyse existing designs, and providing the HRI community with new tools to enrich the ways social robots communicate with humans through sound.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12369-022-00891-0.

Acknowledgements The authors would like to thank Ben Gabaldon, Connor Moore, and Jeshua Whitaker for providing invaluable insight into their design processes.

Funding This research was supported by the University of New South Wales through a Scientia PhD scholarship. Open Access funding enabled and organized by CAUL and its Member Institutions.

Data Availability All data generated or analysed during this study are included in this published article and its supplementary information files. The supplementary data comprise transcripts of the three interviews with the sound designers. The transcripts have been edited for clarity.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The interview procedure was approved by the ethics board of the University of New South Wales. HC No: HC200314.

Consent for publication Consent for publishing the interviews and credit the interviewees was given by the interviewees in writing

Consent to participate Consent to participate was given by the interviewees in writing.

References

1. Aarestrup M, Jensen LC, Fischer K (2015) The sound makes the greeting: Interpersonal functions of intonation in human-robot interaction. In: 2015 AAAI spring symposium series
2. Audiokinetic: Wwise middleware game audio delivery system. https://audiokinetic.com/
3. Aylett MP, Sutton, SJ, Vazquez-Alvarez, Y (2019) The right kind of unnatural: designing a robot voice. In: Proceedings of the 1st international conference on conversational user interfaces, pp 1–2
4. Bramas B, Kim YM, Kwon DS (2008) Design of a sound system to increase emotional expression impact in human-robot interaction. In: 2008 international conference on control, automation and systems., pp 2732–2737. IEEE
5. Braun V, Clarke V (2006) Using thematic analysis in psychology. Qual Res Psychol 3(2):77–101. https://doi.org/10.1191/1478088706qp063oa
6. Breazeal C, Dautenhahn K, Kanda T (2016) Social Robotics. In: Siciliano B, Khatib O (eds) Springer handbook of robotics. Springer International Publishing, Cham, pp 1935–1972. https://doi.org/10.1007/978-3-319-32552-1_72
7. Bretan M, Weinberg G (2016) A survey of robotic musicianship. Commun ACM 59(5):100–109
8. Brock DP, Martinson E (2006) Using the Concept of Auditory Perspective Taking to Improve Robotic Speech Presentations for Individual Human Listeners. In: AAAI fall symposium: aurally informed performanc , pp 11–15
9. Cambre J, Kulkarni C (2019) One voice fits all? social implications and research challenges of designing voices for smart devices. Proc ACM Human-Comput Int 3(CSCW):1–19
10. Carpenter J (2013) Just doesn't look right: exploring the impact of humanoid robot integration into explosive ordnance disposal teams. In: Handbook of research on technoself: Identity in a technological society, pp 609–636. IGI Global
11. Cha E, Fitter NT, Kim Y, Fong T, Matarić MJ (2018) Effects of robot sound on auditory localization in human-robot collaboration. In: Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction - HRI '18, pp 434–442. ACM Press, Chicago, IL, USA https://doi.org/10.1145/3171221.3171285. http://dl.acm.org/citation.cfm?doid=3171221.3171285
12. Chang RCS, Lu HP, Yang P (2018) Stereotypes or golden rules? Exploring likable voice traits of social robots as active aging companions for tech-savvy baby boomers in Taiwan. Comput Human Behav 84:194–210. https://doi.org/10.1016/j.chb.2018.02.025
13. Cicconet M, Bretan M, Weinberg G (2013) Human-robot percussion ensemble: Anticipation on the basis of visual cues. IEEE Robot Autom Mag 20(4):105–110
14. Cohen AJ (1999) Functions of music in multimedia: A cognitive approach. In: Yi SW (ed) Music, mind, and science. Seoul National University Press, Seoul, Korea, pp 40–68
15. Collins K et al (2008) Game sound: an introduction to the history, theory, and practice of video game music and sound design. MIT Press, Cambridge
16. Dahl L, Bellona J, Bai L, LaViers A (2017) Data-driven design of sound for enhancing the perception of expressive robotic movement. In: Proceedings of the 4th international conference on movement computing - MOCO '17, pp 1–8. ACM Press, London, United Kingdom. https://doi.org/10.1145/3077981.3078047.http://dl.acm.org/citation.cfm?doid=3077981.3078047
17. Eyssel F, Kuchenbrandt D, Bobinger S (2012) 'If You Sound Like Me, You Must Be More Human': On the Interplay of Robot and User Features on Human- Robot Acceptance and Anthropomorphism p. 2
18. Fischer K, Niebuhr O, Jensen LC, Bodenhagen L (2019) Speech melody matters-how robots profit from using charismatic speech. ACM Trans Human-Robot Inter (THRI) 9(1):1–21
19. Frid E, Bresin R (2022) Perceptual Evaluation of blended sonification of mechanical robot sounds produced by emotionally expressive gestures: augmenting consequential sounds to improve non-verbal robot communication. Int J Soc Robot 14(2):357–372
20. Frid E, Bresin R, Alexanderson S (2018) Perception of mechanical sounds inherent to expressive gestures of a nao robot-implications for movement sonification of humanoids. In: Sound and music computing
21. Hoffman G, Vanunu K (2013) Effects of robotic companionship on music enjoyment and agent perception. In: 2013 8th ACM/IEEE international conference on human-robot interaction (HRI), pp 317–324. IEEE, Tokyo, Japan. https://doi.org/10.1109/HRI.2013.6483605. http://ieeexplore.ieee.org/document/6483605/
22. Hug D (2010). Investigating Narrative and Performative Sound Design Strategies for Interactive Commodities. In: Ystad, S., Aramaki, M., Kronland-Martinet, R., Jensen, K. (eds) Auditory Display. CMMR ICAD 2009. Lecture Notes in Computer Science, vol 5954. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-12439-6_2
23. Hug D, Misdariis N (2011) Towards a conceptual framework to integrate designerly and scientific sound design methods. In: Proceedings of the 6th audio mostly conference on a conference on interaction with sound - AM '11, pp 23–30. ACM Press, Coimbra, Portugal. https://doi.org/10.1145/2095667.2095671.http://dl.acm.org/citation.cfm?doid=2095667.2095671
24. Jee ES, Park SY, Kim CH, Kobayashi H (2009) Composition of musical sound to express robot's emotion with intensity and synchronized expression with robot's behavior. In: RO-MAN 2009 - The 18th IEEE international symposium on robot and human interactive communication, pp 369–374. IEEE, Toyama, Japan. https://doi.org/10.1109/ROMAN.2009.5326258.http://ieeexplore.ieee.org/document/5326258/
25. Komatsu T, Kobayashi K, Yamada S, Funakoshi K, Nakano M (2018) Vibrational artificial subtle expressions: conveying system's confidence level to users by means of smartphone vibration. In: Proceedings of the 2018 CHI conference on human factors in computing systems - CHI '18, pp. 1–9. ACM Press, Montreal QC, Canada. https://doi.org/10.1145/3173574.3174052.http://dl.acm.org/citation.cfm?doid=3173574.3174052
26. Latupeirissa AB, Bresin R (2020) Understanding non-verbal sound of humanoid robots in films. In: Workshop on mental models of robots at HRI 2020 in Cambridge, UK
27. Lissa Z (1965) Ästhetik der Filmmusik, vol. 73. Henschel, Leipzig, Germany . https://www.worldcat.org/title/asthetik-der-filmmusik/oclc/9898626
28. Löffler D, Schmidt N, Tscharn R (2018) Multimodal expression of artificial emotion in social robots using color, motion and sound. In: Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction - HRI '18, pp 334–343. ACM Press, Chicago, IL, USA . https://doi.org/10.1145/3171221.3171261. http://dl.acm.org/citation.cfm?doid=3171221.3171261
29. McGinn C, Torre I (2019) Can you tell the robot by the voice? An exploratory study on the role of voice in the perception of robots. In: 2019 14th ACM/IEEE international conference on human-robot interaction (HRI), pp 211–221. IEEE
30. Moore D, Dahl T, Varela P, Ju W, Næs T, Berget I (2019) Unintended consonances: methods to understand robot motor sound perception. In: Proceedings of the 2019 CHI conference on human factors in computing systems - CHI '19, pp 1–12. ACM Press, Glasgow, Scotland Uk. https://doi.org/10.1145/3290605.3300730.http://dl.acm.org/citation.cfm?doid=3290605.3300730

31. Moore D, Tennent H, Martelaro N, Ju W (2017) Making noise intentional: a study of servo sound perception. In: Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction - HRI '17, pp 12–21. ACM Press, Vienna, Austria. https://doi.org/10.1145/2909824.3020238.http://dl.acm.org/citation.cfm?doid=2909824.3020238
32. Moore R (2017) Appropriate voices for artefacts: some key insights
33. Moore RK, Mitchinson B (2017) A biomimetic vocalisation system for MiRo. In: Conference on biomimetic and biohybrid systems, pp 363–374. Springer
34. Murch W (2005) Dense clarity - clear density. Trans Rev 5(1):7–23
35. Nakagawa K, Shiomi M, Shinozawa K, Matsumura R, Ishiguro H, Hagita N (2013) Effect of robot's whispering behavior on people's motivation. Int J Soc Robot 5(1):5–16
36. Otsuka T, Nakadai K, Takahashi T, Komatani K, Ogata T, Okuno HG (2009) Voice quality manipulation for humanoid robots consistent with their head movements. In: 2009 9th IEEE-RAS international conference on humanoid robots, pp 405–410. IEEE, Paris, France international conference on humanoid robots, pp 405–410. IEEE, Paris, France. https://doi.org/10.1109/ICHR.2009.5379569. http://ieeexplore.ieee.org/document/5379569/
37. Panariello C, Sköd M, Frid E, Bresin R (2019) From vocal-sketching to sound models by means of a sound-based musical transcription system. In: Proceedings of the sound and music computing conference (SMC)
38. Pelikan HR, Broth M, Keevallik L (2020) " Are You Sad, Cozmo?" How humans make sense of a home Robot's emotion displays. In: Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction, pp 461–470
39. Read R (2014) A study of non-linguistic utterances for social human-robot interaction. Ph.D. Thesis
40. Read R, Belpaeme T (2014) Situational context directs how people affectively interpret robotic non-linguistic utterances. In: Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14, pp 41–48. ACM Press, Bielefeld, Germany. https://doi.org/10.1145/2559636.2559680.http://dl.acm.org/citation.cfm?doid=2559636.2559680
41. Read R, Belpaeme T (2016) People interpret robotic non-linguistic utterances categorically. Int J Soc Robot 8(1):31–50. https://doi.org/10.1007/s12369-015-0304-0
42. Robinson FA, Velonaki M, Bown O (2021) Smooth operator: tuning robot perception through artificial movement sound. In: Proceedings of the 2021 ACM/IEEE international conference on human-robot interaction, pp 53–62
43. Savery R (2021) Machine learning driven musical improvisation for mechanomorphic human-robot interaction. In: Companion of the 2021 ACM/IEEE international conference on human-robot interaction, pp 559–561
44. Savery R, Zahray L, Weinberg G (2020) Emotional musical prosody for the enhancement of trust in robotic arm communication. arXiv preprint arXiv:2009.09048
45. Schwenk M, Arras KO (2014) 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, pp 161–167. IEEE, Edinburgh, UK. https://doi.org/10.1109/ROMAN.2014.6926247.http://ieeexplore.ieee.org/document/6926247/
46. Singer E, Feddersen J, Redmon C, Bowen B (2004) LEMUR's musical robots. In: Proceedings of the 2004 conference on new interfaces for musical expression, pp 181–184
47. Solis J, Chida K, Isoda S, Suefuji K, Arino C, Takanishi A (2005) The anthropomorphic flutist robot WF-4R: from mechanical to perceptual improvements. In: 2005 IEEE/RSJ international conference on intelligent robots and systems, pp 64–69. IEEE
48. Tennent H, Moore D, Jung M, Ju W (2017) Good vibrations: How consequential sounds affect perception of robotic arms. (2017) 26th IEEE international symposium on robot and human interactive communication (RO-MAN. IEEE, Lisbon, pp 928–935
49. Thiessen R, Rea DJ, Garcha DS, Cheng C, Young JE (2019) Infrasound for HRI: a robot using low-frequency vibrations to impact how people perceive its actions. In: 2019 14th ACM/IEEE international conference on human-robot interaction (HRI), pp 11–18. IEEE
50. Tonkin M, Vitale J, Herse S, Williams MA, Judge W, Wang X (2018) Design methodology for the ux of hri: A field study of a commercial social robot at an airport. In: Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction, pp 407–415
51. Trovato G, Paredes R, Balvin J, Cuellar F, Thomsen NB, Bech S, Tan ZH (2018) The sound or silence: investigating the influence of robot noise on proxemics. In, (2018) 27th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 713–718. IEEE, Nanjing. https://doi.org/10.1109/ROMAN.2018.8525795. https://ieeexplore.ieee.org/document/8525795/
52. Walters ML, Syrdal DS, Koay KL, Dautenhahn K, te Boekhorst R (2008) Human approach distances to a mechanical-looking robot with different robot voice styles. In: RO-MAN 2008 - the 17th ieee international symposium on robot and human interactive communication, pp 707–712. IEEE, Munich, Germany. https://doi.org/10.1109/ROMAN.2008.4600750.http://ieeexplore.ieee.org/document/4600750/
53. Wingstedt J (2004) Narrative functions of film music in a relational perspective. In: ISME 2004, 26th international society for music education world conference, 11-16 July 2004, Tenerife, Spain. International Society for Music Education
54. Wolford J, Gabaldon B, Rivas J, Min B (2019) Condition-based robot audio techniques. Google Patents
55. Yilmazyildiz S, Read R, Belpeame T, Verhelst W (2016) Review of semantic-free utterances in social human-robot interaction. Int J Human-Comput Int 32(1):63–85. https://doi.org/10.1080/10447318.2015.1093856
56. Zhang A, Malhotra M, Matsuoka Y (2011) Musical piano performance by the ACT Hand. In: 2011 IEEE international conference on robotics and automation, pp 3536–3541. IEEE
57. Zhang R, Jeon M, Park CH, Howard A (2015) Robotic sonification for promoting emotional and social interactions of children with ASD. In: Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction extended abstracts - HRI'15 Extended Abstracts, pp. 111–112. ACM Press, Portland, Oregon, USA. https://doi.org/10.1145/2701973.2702033.http://dl.acm.org/citation.cfm?doid=2701973.2702033

Frederic Anthony Robinson a german-born musician, audio designer and researcher passionate about the exploration of technologies that change they way we listen to and interact with sound. He has designed and realised interactive multi-channel compositions and soundscapes for large-scale media installations and exhibitions across the globe, including: National Museum of Qatar, Doha; Hyundai Motorstudio Goyang; Museum of the Future, Dubai; Art Stable at the Royal Palace, Oslo; Swiss Museum of Transport, Lucerne; King Abdulaziz Center for World Culture, Dhahran. As a Scientia PhD Candidate at the UNSW Creative Robotics Lab & Interactive Media Lab, he creates exploratory sound design for robotics and smart environments, looking for ways to create richer interactions between humans and the machines around them. As a researcher at Dolby Laboratories, he prototypes creative applications of next-generation audio technologies.

Ollie Bown is a researcher and maker working with creative technologies. He comes from a highly diverse academic background spanning social anthropology, evolutionary and adaptive systems, music informatics and interaction design, with a parallel career in electronic music and digital art spanning over 15 years. He is interested in how artists, designers and musicians can use advanced computing technologies to produce complex creative works. His current active research areas include media multiplicities, musical metacreation, the theories and methodologies of computational creativity, new interfaces for musical expression, and multi-agent models of social creativity. He is an associate professor at the Faculty of Arts, Design & Architecture, University of New South Wales.

Mari Velonaki 's research is situated in the multi-disciplinary field of Social Robotics. Her approach to Social Robotics' research has been informed by aesthetics and design principles that stem from the theory and practice of Interactive Media Art. Velonaki has made significant contributions in the areas of Social Robotics, Media Art and Human-Machine Interface Design. Her career outputs across these fields are extensive. Velonaki began working as a media artist/researcher in the field of responsive environments and interactive interface design in 1997. She pioneered experimental interfaces that incorporate movement, speech, touch, breath, electrostatic charge, artificial vision and robotics, allowing for the development of haptic and immersive relationships between participants and interactive agents. She is the recipient of several competitive grants, including ARC Discovery, Linkage, LIEF an ARC Fellowship, an Australia Council of the Arts, Visual Arts Fellowship, Australia-Japan Foundation, Fuji Xerox Innovation, AOARD. Velonaki is a Professor of Social Robotics at Art & Design, UNSW. She is the founder and director of the Creative Robotics Lab (Art & Design UNSW) and the founder and director of the National Facility for Human Robot Interaction Research (UNSW, USYD, UTS, St Vincent's Hospital). Mari's robots and interactive installations have been exhibited worldwide, including: Victoria & Albert Museum, London; National Art Museum Beijing; Gyeonggi Museum of Modern Art, Korea; Aros Aarhus Museum of Modern Art, Denmark; Wood Street Galleries, Pittsburgh; Millennium Museum - Beijing Biennale of Electronic Arts; Ars Electronica, Linz; European Media Arts Festival, Osnabruck; ZENDAI Museum of Modern Art, Shanghai; Art Gallery of NSW, Sydney, Museum of Contemporary Arts, Sydney; Conde Duque Museum, Madrid.