Robot Sound in Distributed Audio Environments

Frederic Anthony Robinson*†

Mari Velonaki*

Oliver Bown†

Creative Robotics Lab* & Interactive Media Lab†

University of New South Wales, Sydney NSW

frederic.robinson@unsw.edu.au

1 Introduction

While a robot's ability to listen spatially has been an active field of research for a while, robots emitting spatial sound have received significantly less attention, as social robots are typically treated as a single sound source with access to a single loudspeaker. However, with the growing availability of audio-enabled networked devices in the home, one may imagine a future where (i) these devices become coordinated networks of distributed speakers emitting spatial sound throughout the home, and where (ii) social robots will be able to emit sound across these systems. This raises interesting questions around the role and benefit of spatial cues in a domestic robot's communication.

From a sound designer's perspective, a robot that has access to more than one speaker can begin to emit spatial sound and thereby draw from an extended pool of perceptual effects when communicating with humans. Spatial sound – positioning different sounds at different locations around the listener – creates a more immersive experience, enabling the creation of unreal, yet convincing sound environments [35]. Experiencing ambient soundscapes in this manner, for example, can make people feel calm or vibrant [12], more or less safe [37], or even affect how they recover in health care facilities [29]. More generally, reproducing sound in higher spatial fidelity has shown to substantially contribute to perceived audio quality [36].

From a human-robot interaction (HRI) perspective, robot sound being able to *leave* a robot's body has an additional implication. Using an anthropomorphism, the robot may be considered a *body* (hardware) and a *mind* (software) that can be disembodied. While many of the robot mind's capabilities are shared with other entities like voice assistants, the robot's physical body is a distinctive characteristic. The unique implications of a robot's physical presence have been thoroughly explored by human-robot interaction (HRI) researchers. See, for example, the work by Kiesler and colleagues, who compared how humans interact with a physical robot and a robot on a screen [24]. Among other things, they found that participants were simply more engaged with the physical robot. More recently, Li surveyed a large number of experimental works investigating this phenomenon and concluded that robots are perceived more positively and are more persuasive when physically present [28]. Sound, and particularly sound distribution, plays an interesting role in this mind-body dichotomy, as robot sound can be attributed to a physical robot when it is emitted by the same, but it may well be attributed to some other source when it is perceived from somewhere else in the environment. As a result, distributed sound may turn this *physical presence* from the inherent characteristic of a social robot into a temporary state that can be changed at will. This poses interesting questions around a robot's identity, such as "How is a robot's perceived identity affected when it is no longer associated with its physical structure, and might this be desirable in certain interaction scenarios?"

Re-embodiment, the transfer of one single social intelligence between several physical robots, is a design paradigm that has recently received attention among HRI researchers. Luria and colleagues, for example, had researchers and designers explore the "vast design space of social presence flexibility" in a User Enactments study [30, p. 635]. Reig and colleagues explored user acceptance of a robot personality moving between multiple robot bodies in a service context [32]. Both studies reported that participants showed general acceptance of the practice. While these works consider the transfer of a specific type of sound, the voice, between multiple specific locations, robot bodies, it is not far-fetched to consider the distribution of robot sound across arbitrary locations. The latter has, in fact, been explored by Iravantchi and colleagues, who used an ultrasonic array, a highly directional loudspeaker, to project speech onto arbitrary objects in a living space, giving a "voice to everyday objects" [23, p. 376].

In the case where no such specialised equipment is available, one might ask where these speakers in environment should come from. Can we expect to have loudspeakers present in a robot's environment, and can we expect a robot to have access to these loudspeakers? Strong indicators that both of these prerequisites will eventually be met can be found in current developments in the internet of things (IoT) ecosystem. A growing number of hardware devices with networking capabilities permeates public space, work places, and the home [3]. Many of these devices have loudspeakers built into them, and efforts have begun to use these capabilities to create distributed audio playback systems. Electronics manufacturer LG, for example, patented an implementation of smart connected distributed audio systems in the home using miscellaneous networked devices such as TVs, fridges, and, interestingly, also robotic vacuum cleaners. Their implementation utilised approximate position tracking and streaming via the 5G network [25]. It is reasonable to expect that robots will eventually be able to make use of these systems, and the various challenges and opportunities surrounding this new design space are, in our view, worthy of exploration. In light of these considerations, this chapter aims to map out a design space for spatial sound in human-robot interaction by asking the following question:

"How might a robot's auditory communication be enhanced by being distributed across loudspeakers in the environment?"

To explore this question, we first conduct interviews with researchers and practitioners with distributed audio and sound installation experience, identifying design themes for applying interactive sound installation techniques in the context of human-robot interaction. These insights, combined with the authors' own expertise in interactive immersive audio environments, then inform the creation of a virtual distributed robot sound prototype. This process includes the ideation, and realisation of spatial robot sound. After presenting the prototype and its design process, we reflect on lessons learnt and propose a generalised design framework for spatial robot sound. Rather than attempting to capture all of the experimentation presented in this chapter, the framework provides a pragmatic formalisation of what we argue to be key aspects of spatial robot sound.

2 Research Context

In order to provide additional context for this research – how human-robot interactions might benefit from distributed robot sound – two questions need to be answered, "What is the current research on distributed sound in the home?" and "What is the state of current research around distributed sound in human-robot interaction?"

2.1 Distributed Sound in the Home

A discussion of the technologies around networked audio systems in the home is beyond the scope of this chapter. It is, however, worthwhile to consider current research around the design challenges faced in this context. A notable concept in this space is what Francombe et al. coined *media device orchestration* (MDO) to describe the "concept of utilizing any available devices for optimum reproduction of a media experience" [19, p. 3]. Their motivation is partly based on user studies suggesting that more general spatial audio experiences such as envelopment and image depth are more relevant to the user than precise localisation [31]. This potentially makes MDO a viable alternative to precisely calibrated loudspeaker systems that can reproduce spatial audio in high resolutions and may enable novel and improved multimedia experiences. Listening tests utilising MDO showed, that low-channel count sound material could be successfully augmented by being distributed across various devices, such as phones and tablets, traditional stereo systems, and TVs [43]. An early audio drama by the BBC which utilised the technology was well received among listeners [18]. More experiences have been created since then. Discussing the concrete implementation of this type of audio content into a broad range of loudspeaker setups in homes, Francombe et al. state that "it is important to understand both the required metadata and rendering methods to best select devices for different object types and audio signal features" [20, p. 424]. Defining a media format to enable mass distribution remains an ongoing challenge, raising questions concerning both content creation as well as the role of the end-user, who in the current implementation has to report the device positions manually [21]. The researchers have created a toolkit that allows designers to create their own interactive spatial sound experiences [4].

2.2 Distributed Sound in HRI

When considering the question how robots might fit into such an environment, it should be noted that robots have long been applied within smart environments in health and aged care contexts [10]. Bordignon et al. state, that the areas of robotics and ambient intelligence are "converging toward the vision of smart robotic environments" [8, p. 3101], and so far, various research in this domain has focused on establishing the required technological and conceptual frameworks [16, 42, 11, 41]. Sound has, to our knowledge, seen little consideration in this context, and research on robot sound with a spatial component is generally sparse. The context it has appeared in thus far is mainly that of localization - using the spatial cues embedded in a robot's sound to make it easier to localise by a human sharing the same space. The benefit of this is improved spatial coordination and lower chances of collisions [17]. In a study investigating the effect of sound on human-robot collaboration, Cha and colleagues continuously sonified robot movement and found that specific sound profiles made the robot easier to localise for the human [14]. In this work the robot sound source and the physical location of the robot were treated as one and the same. A disconnect between the two, however, was discussed by industry sound designers interviewed in a previous study of ours [33]. Toy robot Cozmo, for example, has a complete parallel music system with orchestral recordings emitted by a companion app on a mobile device. Robot companion Vector had the opportunity to emit sound via surrounding smart devices, but its designers decided against it in order to not distract from its pet-like character. Due to its speaker characteristics, the voice of social robot Jibo sometimes appears to come from the walls around the robot, which was described by its sound designer as a design flaw.

3 Expert Interviews

Due to the exploratory nature of this research, we chose to collect qualitative interviews on the participants' design practice in order to gather data that is both rich and detailed [39, 2]. The expert interviews were motivated by two key questions, "How is the medium of sound applied in immersive environments?", and "How might practices from that domain be translated into the HRI context?" Themes emerging from the analysis of these interviews should then provide a series of design considerations around the use of interactive sound in immersive environments, which could be used to inform the design of a spatial robot sound prototype. The semi-structured interviews were loosely based on a shared set of questions to ensure all relevant areas are covered [6]. However, the questions were adjusted to target relevant projects from the interviewees respective portfolios. Participants were also encouraged to add any information they felt was not covered by the questions to reduce the questions' role in shaping the responses [22]. This approach allowed us to (i) ask detailed questions on the participants' design process for existing works, and (ii) encourage them to envision how these processes might be applied to the context envisioned in this paper: spatial robot sound for the home. It should be noted that interviewees were not familiar with this new context. It was therefore the interviewer's task to encourage participants to apply experiences from prior work to this new context without preempting their own opinions on the subject [27]. The interviews took between 45 and 75 minutes and were conducted remotely via video call. The resulting six hours of audio recordings were then transcribed.

3.1 Thematic Analysis

To analyse the transcripts, we used thematic analysis, an inductive technique to identify and cluster themes from interview data [9]. This process involves the careful examination of qualitative data to identify patterns and extract meaning [26]. This was done in several stages. In a first step, we became familiar with the data by reading the transcripts to understand the overall message behind participants' answers. In a second step, the transcripts were codified, meaning important features of the data were marked with succinct tags. The codes were then separated into categories to identify broader patterns in the data. Finally, we generated themes to describe the categories. To provide additional context with these themes, we additionally quote participants, as recommended by Bechhofer and Paterson [5].

3.2 Participants

The participants include designers and researchers working with interactive immersive sound environments. This can range from technical work such as speaker planning and system design, through conceptual design, to sound design practice. We chose experts with this background because of the common ground between their disciplines and the focus of this paper: using interactive spatial sound to create rich and engaging experiences for the listener.

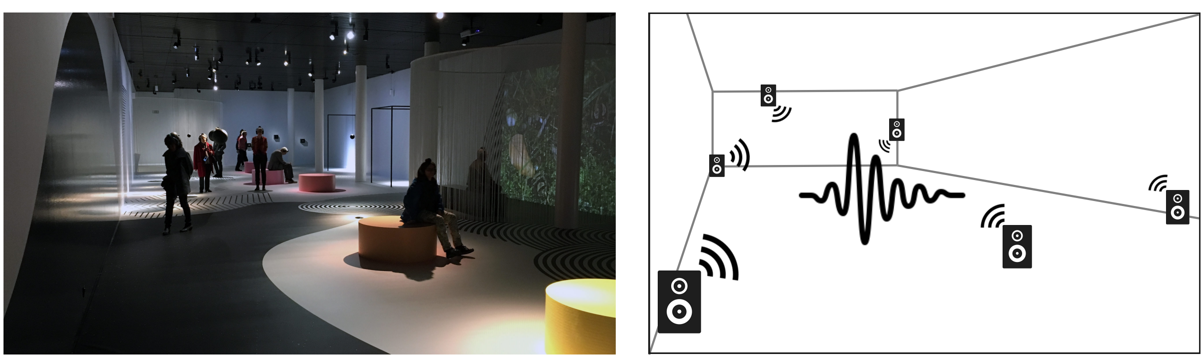

Figure 1: Left: An exhibition space. Right: Distributed sound sources create an immersive environment.

An example of this kind of work can be found in site-specific media installations, where designers use distributed and interactive audio throughout museum spaces to create themed, coherent sound and music environments that reinforce core messages of the exhibition [15]. An environment with distributed sound sources is illustrated in Figure 1. Several of the interview participants had previously worked on projects like these, which additionally touched on themes relevant to the HRI context. Among others, the projects included (i) a large, cloud-like audiovisual installation that used interactive sound to convey notions of intelligence and animacy, (ii) an exhibition about the history of automata and robots, featuring sound for the specific machines exhibited, as well as enveloping soundscapes, and (iii) a museum that had a bodiless artificial intelligence follow visitors through the exhibition area.

3.3 Themes

Five main themes emerged during analysis of the expert interviews. They are *sonic identity and robot fiction*, *functions of sound*, *roles and affordances of interactive sound*, *roles and affordances of distributed sound*, and *technical considerations*, and are shown in Figure 2.

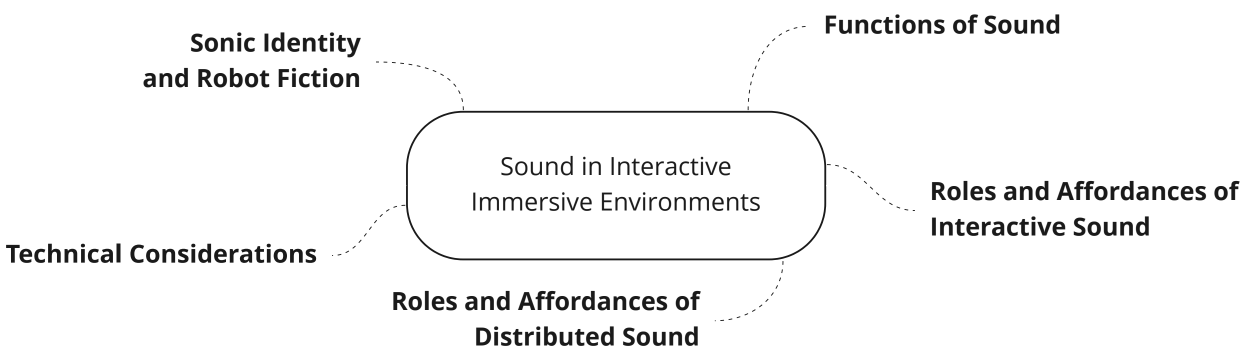

Figure 2: Five themes emerging from expert interviews with researchers and designers working with sound in interactive immersive environments.

3.3.1 Sonic Identity and Robot Fiction

A key notion mentioned by several of the experts was the idea of creating a coherent, consistent, and holistic sonic identity, both when discussing immersive sound environments and when speculating on the embedding of robotic agents into them. Motivations for this are not only a more conceptually refined design, but also potential branding considerations. This idea of sonic identity matches well with the notion of robot fiction, mentioned by robot sound designers working towards creating a believable robotic character [33]. Sonic identities are created by putting all emitted sound in relation to each other, or, in the words of one expert, by “making them harmonize.”

Often, references to existing sound sources are used to achieve this. Creating sound for an exhibition on robotics, one expert drew from recordings of historical industrial equipment. Creating a soundscape whose particular timbre and internal rhythm was defined by this idiosyncratic but real sound source was “immediately enhancing the visitors’ projection.” In their words: “You know this is not real. This is quite clearly artificial, but you want to believe in it. You want it to be real.” This referencing was also said to potentially constrain the designer’s creative freedom. One expert noted how building sonic identities around rigid references can provide a “narrow” design space. As an example, they mentioned the sound of an electric vehicle, which can be a digital recreation of a combustion engine, or it can be "invented from the ground up." A middle ground was discussed by a third expert, who was tasked with creating affective sound for an installation containing a responsive structure meant to be perceived as animate. They recorded the human voice non-verbally expressing various emotional states, and then processed them to turn them into utterances of an artificial character. In this instance, loosely referencing the human voice provided an "organic" source material, while the processing of the sound material made it part of a non-human sonic identity. As a result, the responsive structure was perceived as an "independent thing that is [...] not just trying to imitate."

Opinions differed on how much interactivity or personalization should be allowed to work against or even break sonic identities. One expert noted that a user's influence on the soundscape should be limited in a way that does not allow for the fiction to be broken. The core message of the sound is unchangeable, and interactivity is a thin layer above it. Another expert felt the sonic identity of a robot should be able to be easily changed and personalised. According to them, people may want different things out of the same robot model – or even the exact same robot – and sound is a key differentiator here, as it is the only thing that can be significantly transformed at any point during the robot's deployment.

Space is a relevant parameter for playing with the notion of sonic identity, as sound distribution has previously been flagged as a sound characteristic that can impact a robot's fiction. One expert described a robot embedded in a sounding IoT environment as both a self-contained character and an interface through which one could interact with their smart home. Both of these cases imply certain sound distributions. Robots Cozmo and Vector's sound designer Ben Gabaldon stressed how his robots could have emitted sound through smart devices in their environment, but this was avoided specifically to not break the robots' fiction [33].

3.3.2 Functions of Sound

Experts highlighted three general aims in their sound designs: (i) conveying and causing various emotional states and moods, (ii) providing orientation and guiding a listener's attention, and (iii) creating associations to external concepts, environments or things. Those who worked on robotics-related projects specifically highlighted the use of associations to convey animacy – "how can you deceive someone into thinking that this is more lifelike than it is?" – and to differentiate the various elements that come together to form a robotic agent. One expert suggested using recordings of real sound sources to convey information about the robot body, and purely synthetic sound and musical patterns to convey information about the robotic mind, such as data flow, networks and connections between several robots.

Another expert proposed five categories of sound to be used in a robot's communication: speech, sounds of the body, alert sounds, sounds used for orientation, and moods and atmospheres. For the latter, they suggested the use of music for conveying mood, and more realistic ambient soundscapes for illustration, noting that both of these are a promising new way of communicating that could be unique to the machines, as neither are part of a human's communication. Several experts mentioned the potential of combining these two by embedding musical material into ambient soundscapes to create abstract but emotional environments.

3.3.3 Roles and Affordances of Interactive Sound

In the context of this chapter, *interactive sound* describes sound that is in some way mapped to states and actions of a human, robot, or the environment. The experts discussed various ways to make sound responsive, both in simple and complex ways. Data that could be used to affect sound includes the number of people present, weather data, time of day, distance to humans, and temperature. Sound parameters to be affected could include volume, timbre, choice of sound material, and audio effect processing. One expert described how these mappings could be combined to build sound environments that are perceived as a "basic intelligence." They also mentioned, however, that one can "take a very simple piece of sensor information like distance [and] use it to communicate a lot." In an application of this, they used distance information to make a responsive structure convey increasingly negative and agitated emotional states as it was being approached, communicating not only that it is aware of its surroundings but also hinting at underlying personality traits. In another interactive installation, they distinguished between three different types of interactivity, (i) *sharing*, where the installation communicates but does not respond to the listener, (ii) *subtle interactivity*, where the listener feels like they affect the soundscape but do not know how specifically, and (iii) *play*, where there is an explicit interaction between machine and human.

When asked about specific generative techniques, one expert described the use of several mapping layers to first translate arbitrary robot sensor data into parameters that are relevant to the interaction, before then using those parameters to control sound output. For example, we may not be interested in a robot wheel's rotations per minute, but observing the motion of all wheels over time, and combining them with information about the surface it is moving on, may then be translated to a parameter called "exertion." This information can then be mapped to various sound characteristics to convey the notion of strain. As a result, a simple model of a biological muscle is inserted between sensor data and sound mappings, essentially allowing the designer to define interactive behaviour in two stages, first translating raw robot data into meaningful descriptors of robot behaviour or interaction scenario, and then translating these descriptors into sound. It also means that the first half of this responsive sound system is independent of the actual sound generation methods being used. The expert notes that this allows the designer to freely choose between different approaches like real-time synthesis, or sample-based playback.

3.3.4 Roles and Affordances of Distributed Sound

The potency of immersive sound to create a range of powerful listening experiences and was a recurring theme among experts, who highlighted how individual sound sources from locations around the listener can together form rich and believable environments. The ability to place sound at different locations around the listener allows for a "huge level of complexity" to emerge. One expert mentioned how the use of "tangential sounds" played through loudspeakers across a space can "give this sense of detail that [people] are used to in a real environment," allowing the listener to "project a lot further." Another expert mentioned how distributing sound across a space not only allows for the creation of detailed environments but also for the creation of clear distinction and separation. Referencing a prior project, they described how individual voices of a recorded string quartet were separated spatially, thereby making it easier for the listener to focus on individual instruments.

When asked about the potential of distributed or spatial sound for robot communication, many experts discussed the boundaries between the robot sound and sound in the environment. One described a project in which they placed microphones inside of an animatronic and projected the internal sounds of the machine into the surrounding space. When the animatronic would perform gestures, the sounds associated to these gestures would then be amplified across the room, at times in a dramatic fashion. As a result, the gestures of the machine were staged in a way that made them highly noticeable and relevant to the listener. Another expert described a project that, instead of using the space to accentuate actions by the physical robot, used space to blur the boundaries between robot and its environment. A media exhibit that was meant to appear animate had built-in speakers, as well as speakers across the exhibition space. When the exhibit communicated, the entire space appeared to resonate.

A third expert described the possibility of a robotic sound source temporarily leaving its physical body to travel to locations in the home that the body might not be able to reach in time, or at all. As an example, one might imagine a robot planning to enter the kitchen, and before doing so, sending only its voice onto a loudspeaker in the kitchen to announce its arrival. The expert described this disembodied state as the robot being *schizophonic*, a term coined by composer R. Murray Schaefer, which refers to sound being separated from its original source, a misalignment between the auditory and the visual [38].

3.3.5 Technical Considerations

Experts were asked to elaborate on technical considerations when working with sound across an immersive sound environment, and what they would consider noteworthy when imagining robots in such a context. A common concern was variety in speaker distribution to create depth and give designers access to a high number of diverse spatial locations. One expert illustrated this with an example where loudspeakers distributed across the ceiling may evenly cover an area, but still provide sound from only one general direction, above. In their words, "we don't want the room to be in the ceiling." Another expert noted that with envelopment being a key perceptual attribute, distributions do not necessarily need to resemble standard speaker layouts such as 5.1 to be effective. This differs from the well-established audio research paradigm of localization accuracy, where the quality of a spatial sound experience is often quantified by measuring the smallest noticeable offset of degrees between two sound sources. The expert instead suggested to spatialize sound across broad and clearly distinguished categories, such as close vs distant, point source vs enveloping, and moving vs static.

Focusing on the audio capabilities of the robot itself, one expert highlighted the dispersion characteristics

Figure 3: An overview of the prototype design process.

of a robot's speaker – meaning whether a loudspeaker emits sound in a targeted or omnidirectional fashion. A robot emitting sound in all directions will have better speech intelligibility for people not faced by the speaker, but this might not be desirable when they are not part of the interaction. A robot with a highly directional speaker, in turn, will have to adjust its orientation to face a specific listener that is to be addressed. In their eyes, distributing robot sound to other speakers could then be a way to apply these effects in a targeted and refined way. However, this reliance on using external speakers for communication would also limit functionality in deployment locations that do not have additional speakers in the environment.

4 Distributed Robot Sound Prototype

After presenting the various themes emerging from the expert interviews, we now take key notions from these findings and work them into a design prototype: a distributed robot sound profile. Based on the themes, A) *sonic identity and robot fiction*, B) *functions of sound*, C) *roles and affordances of interactive sound*, D) *roles and affordances of distributed sound*, and E) *technical considerations*, we chose the following design goals:

- Create a holistic robot sound set based on a core fiction (A), which incorporates utterances, music, movement sound, and ambient sound (B).
- Design the robot sound set to take into account sound emitted by both the robot and the surrounding space (D).
- Implement the sound set in a virtual environment, so that spatial distribution and responsive behaviour can be prototyped without being restricted by technical constraints (E).
- Make the sound set responsive to user presence and activity (C).

An excerpt of the resulting prototype is shown in Video Excerpt 1. It showcases the sound of various behaviours of robot mind and body in a virtual environment. It also shows the spatial distribution of sound elements, by placing them around a centered listener. The video has a binaural spatial audio soundtrack which, when listened to with headphones, positions sounds around a virtual listener standing in the center of the room. As a result, sound taking place in the top half of the room is positioned in front of the listener, while any sound in the bottom half is positioned in the back.

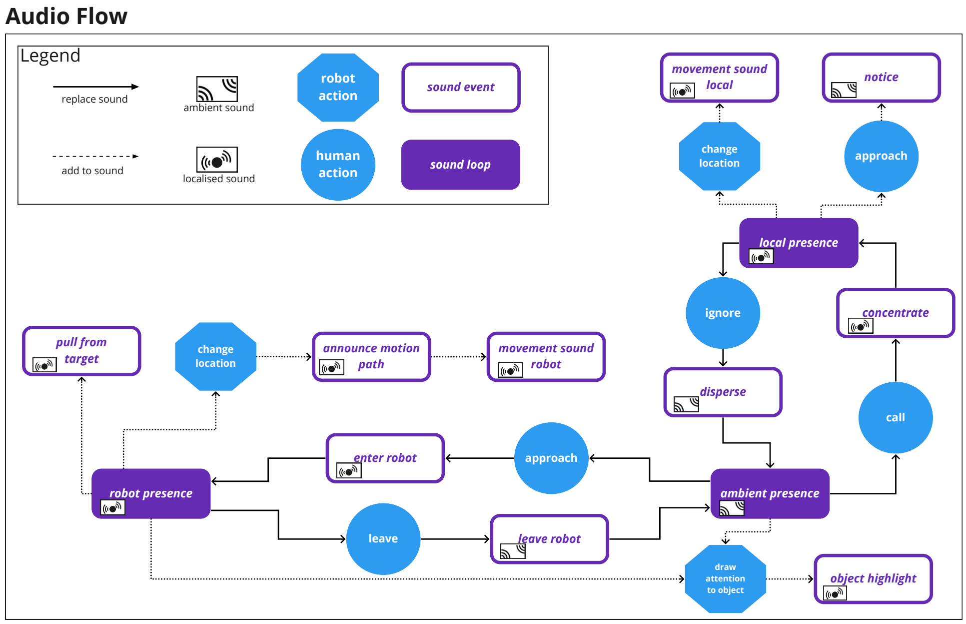

Figure 4: The prototype's Audio Interaction Flow, a flow diagram giving an overview of all sound assets and their interactive relationships.

The following section details the prototyping process and discusses the various design decisions made. Based on a robot sound design process presented in prior work [33], we followed a five-step design process (see Figure 3). It involves (i) the creation of what we call an *Audio Design Board*, (ii) the collection of sound sources making up the *sound identity*, (iii) the creation of a flow diagram we call *Audio Interaction Flow*, (iv) the *production of audio assets*, as well as (v) the final *implementation* of said assets into a virtual robot. The Audio Design Board collects a number of high-level metaphors that informed the robot characteristics, personality, and behaviours we were to design for, and the associations we wanted to evoke. To create the robot's sound identity, we collected sound materials (recordings) and sound sources (synthesis methods), which together would make up how the robot sounds. The identity should sound coherent and consistent and be appropriate for and reflective of the robot's application context and capabilities. The next step was the creation of what we call an Audio Interaction Flow (see Figure 4), a modified user flow diagram that represents the entire robot sound set, and how it responds in different interaction scenarios. This flow diagram then formed the foundation for the asset production, the creation of all individual sound loops and events to be emitted by the robot and its environment. Finally, during implementation, the sounds were reassembled in an interactive audio engine. In the following sections, we would like to specifically address two steps in particular: the Audio Design Board, and the Implementation.

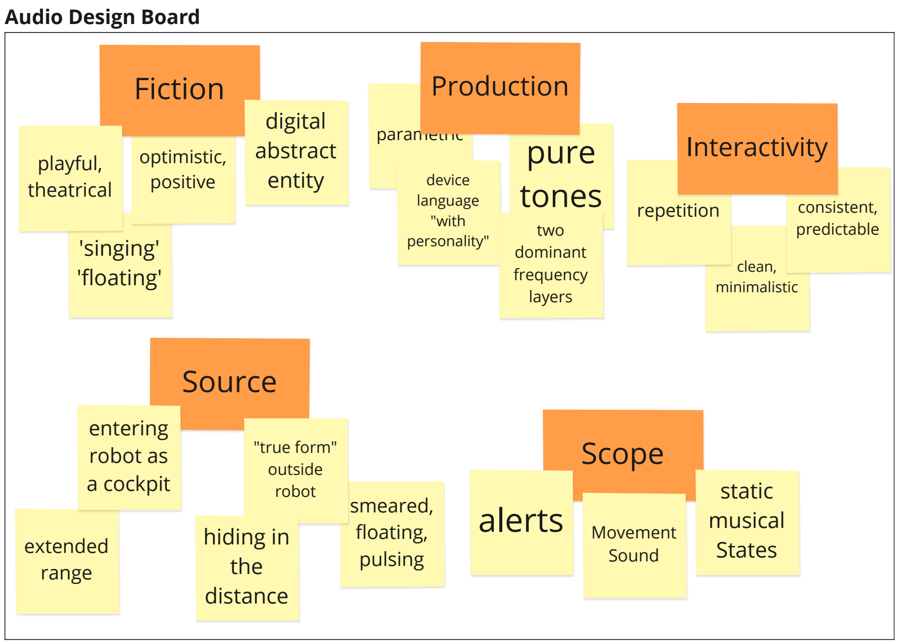

Figure 5: An Audio Design Board bringing together a collection of metaphors and associations, derived from five top-level design themes identified in interviews with robot sound designers [33].

4.1 Audio Design Board

Figure 5 shows the various ideas, metaphors, and associations that formed the foundation of the prototype design. Their high-level categorisation is based on the candidate design principles we identified in previous work [33]: fiction, source, scope, interactivity, and production. The principles provided a framework to assist in creating a complete and internally consistent robot sound set which, in this case, should be emitted both by the robot itself and across its environment.

We chose the **fiction** to be that of a playful and theatrical abstract digital entity. The robot fiction should be technological in nature, and abstract enough to take different forms, such as freely moving across the room, or being embodied as a robotic agent. We also wanted the entity to be expressive, as it would allow us to more freely explore different sound design options instead of working with a more muted sound pallet.

The robot sound **source** should comprise both sound emitted by the physical robot body, as well as any position in the environment. The "true form" of the entity should be apparent when it was freely moving around. Sonically, this should be reflected by sounds with a smeared, floating, and pulsing quality, as if hard to grasp or clearly make out. Entering the robot body would then be reflected in sound through a more grounded, low-frequency range, and more tactile sound with clearer transients and more natural sounding material. However, this range of sounds - from smearing to tactile – should be derived from the same core sound material through editing and processing, giving a coherent core identity to the overall sound pallet.

The **scope** of the robot sound should comprise movement sound, static musical states, and alerts. The latter could also be considered semantic-free utterances, but we chose to stay closer to more traditional UX sound conventions, because the entity was not exclusively embodied, but could also emit sound through smart speakers.

Interactivity refers to the way robot sound is modified through parameters in the environment or through randomness to create a more varied and rich listening experience. For this prototype, we decided to favor the simple playback of single sound files, and not introduce any additional variation. Any robot action would therefore always result in the same respective sound event being played back. We chose this for two reasons. The first was the prior mentioned link to smart devices, which encouraged the use of a UX sound-based language which rarely features this kind of variety. The second reason was that the spatial distribution of sounds in itself introduced extensive variety. Robot body, arbitrary positions across the room, and the whole space itself could all be the emitter of sound. We therefore chose to limit variety within the sound content to reduce complexity. The sound of the robot mind concentrating in the room could therefore happen at any position in the room, but would always be announced by the same sound. Generally, this approach is comparable to that for commercial robot Kuri, which features consistent, repeating sounds that should be unintrusive, clean, and simple [33].

Finally, the **production** of sound assets aimed for a pallet of pure, clean sounds with limited complexity. We chose to use primarily sine tones with some additional harmonics created through the use of FM synthesis, a process in which several sine tones modulate each other to create more complex spectra. A second pool of sound material consisted of high frequency textures created through the granulation of acoustic recordings. This pool was heavily processed to give the sound a digital and artificial quality. As a result, we ended up with two dominant frequency layers, a selection of pitched, pure tones, and a second layer of smeared high-frequency textures. The former would then be used to create alerts in the style of traditional user interface sound, while the latter was used to make sounds easier to localise in the space around the listener – the human ear is better at perceiving positional cues in higher frequencies. We revisited all five themes mentioned above throughout the rest of the design process, both to get inspiration for sound choice and processing, and to ensure that the sounds we created were consisted with the overall design goals established in the beginning.

4.2 Implementation

The interactive behaviour of the prototype can be summarised as follows: The robot sound has three major states, which we describe as *presences*. A *robot presence* is emitted when the robot mind is present in the robot body. In this state, the robot can experience a *pull from a target* location – for example, when being asked to move somewhere – and it can *announce a motion path* which is then subsequently accompanied by a *movement sound*. The second state is the *ambient presence*, in which the robot mind is present and available across the entire space. Leaving and entering the physical robot body are accompanied by a *leave robot* and *enter robot* sound event respectively. When embodied or present across the space, the robot can *highlight an object* in the environment with a sound event. The third state is a *local presence*, during which the robot mind is not embodied, but localised at a specific point in the environment. This state is accessed through a *concentrate* sound event, and left through a *disperse* sound event. When in the local state, the robot mind has its own disembodied movement sound, and can acknowledge events happening around it.

After producing audio assets for all of these states and events, we implemented them into a robot to experience them in context. For this chapter we chose to implement a virtual prototype, rather than deploying a physical robot. This was done for two reasons:

(i) Earlier work co-authored with Albastaki and colleagues demonstrated the value of Virtual Experience Prototypes (VEPs) as tools for prototyping and evaluating robot behaviours, including robot sound [1]. We created a virtual version of a physical robot and embedded it in a virtual environment which resembled a prior deployment location of the physical robot. We found VEPs to be "lightweight in development and deployment" enabling evaluations that are "location-independent with a broad pool of potential participants" [1, p. 84]. We also found that impressions of the virtual robot closely resembled those of the physical robot in an earlier real-world deployment. While remote data collection was not part of our goals for this chapter – a physical deployment of the robot will be presented in future work – we were interested in using the VEP's inherent speed and ease of use for our prototyping workflow, and wanted to gather early impressions of the final outcome ourselves.

(ii) The second reason for virtual deployment was the focus of the research: distributed sound in the robot's environment. With the goal being the exploration of arbitrary sound distributions, we wanted to work in a prototyping environment that supports this. Different from the very site-specific loudspeaker distributions in prior media-installation work – and loudspeaker distributions across a robot's body described in previous work [34] – we wanted to work with abstract sound trajectories that might apply to the arbitrary loudspeaker distributions promised by the current IoT developments in the home (see Section 2.1). A virtual sound environment provided this abstraction.

We therefore embedded a 3D model of robotic artwork Diamandini, discussed in prior work [34], within a custom-made virtual environment created in the Unity game engine. The virtual environment, which is modeled after the University of New South Wales' National Facility for Human-Robot Interaction Research, is shown in Figure 6. Mouse-based interactions with the virtual prototype are shown in Video Excerpt 1. The remainder of this section will describe the various components making up the virtual environment. An overview of the system is shown in Figure 7.

- The virtual environment contains a 3D model of the robot Diamandini and various visual elements that showcase the location of the robot mind at any given time and visualise the distribution of key sound sources. When the robot mind is embodied in the robot body, the model lights up. When the robot mind is present across the space (ambient presence) the entire room is filled with small glowing particles. When the robot mind, and by extension its sound, concentrates at a particular location, these particles concentrate as well.

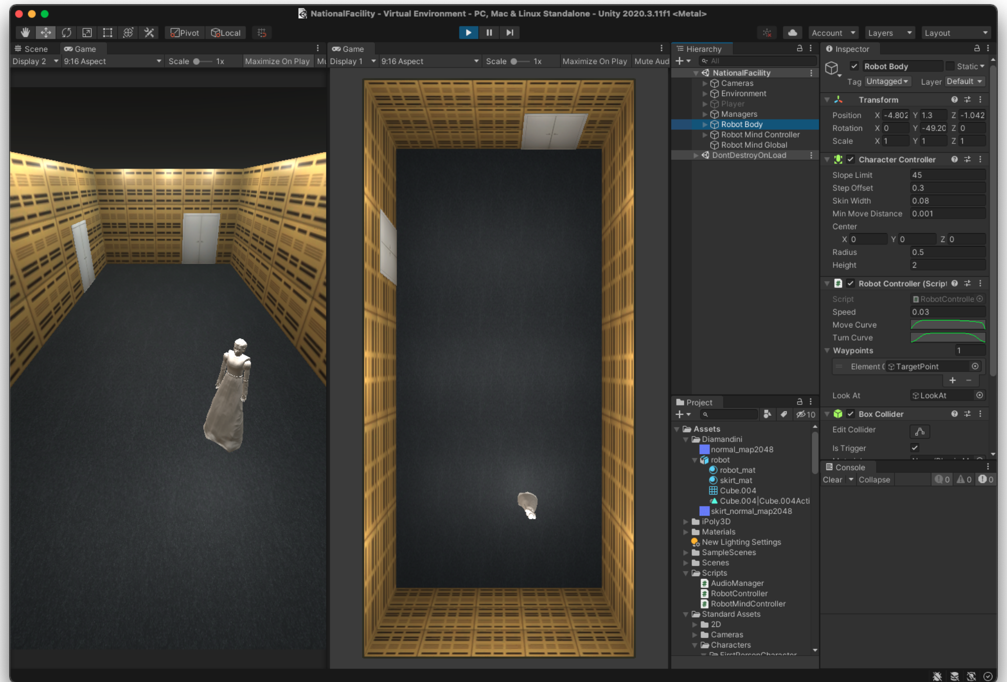

Figure 6: Prototyping spatial robot sound in a virtual environment.

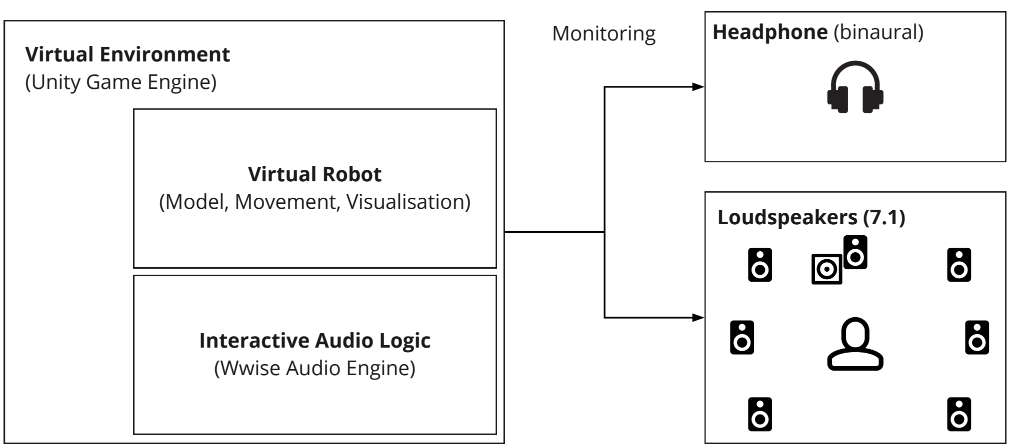

Figure 7: Overview of the virtual environment. Spatial sound can be monitored either through headphones or loudspeakers.

- We programmed the model and the visualisations to respond to mouse and keyboard interactions. Clicking on any location in the space moves the robot to that location (when it is embodied or localised) or makes it concentrate at that location (when it is an ambient presence across the space). Double clicking any location in the space makes the robot mind leave the robot body (when it is embodied) or disperses it (when it is a local presence). Clicking on the robot model makes the robot mind enter the body, and various click and key combinations announce motion paths or highlight objects.
- The soundscape of the virtual environment can be heard via the standard audio output of the game engine. This means playback can be heard via headphones using binaural rendering, which renders a 3D audio environment to two headphone channels. This can be heard in Video Excerpt 1. Audio can also be played back via a 7.1 loudspeaker system in our own, or any other studio with a surround sound system. This is how we worked with the prototype. It also means that running the virtual environment in a room with a surround sound system projects the virtual space into the listening space and thereby provides a way to listen in on how any designed sound set might sound in a home environment.
- We used the Wwise audio engine, embedded within Unity, to specify interactive audio behaviour. One particular benefit of this modular approach is that the robot sound is not hard coded into the virtual environment and can therefore be taken out of the game engine and used in any other context. A design tested and finalised inside the virtual environment can therefore easily be exported and used within an actual physical robot. This is the approach taken for a physical deployment discussed in future work.

5 Findings and Discussion

After presenting the prototype, this final section contains a critical reflection of the design work done in this chapter, and subsequently generalises the findings in form of a design framework for distributed robot sound. Together, these two sections address the question "How might a robot's auditory communication be enhanced by being distributed across loudspeakers in the environment?" Finally, the section discusses limitations and future work related to the research in this chapter.

5.1 Lessons Learnt

5.1.1 Distributed Robot Sound

Needless to say, extending robot sound to any location in the environment results in large number of design possibilities. If we, for example, assume a rough spatial resolution of 1 meter – meaning we can move a sound across a space in 1 meter increments – this would give us around 30 possible locations for any sound emitted in a standard-size living room. Some of these locations are functional, such as highlighting relevant objects in the environment. Many others could serve a more aesthetic function, such as enveloping the listener or adding an additional creative dimension to sound events. A key parameter in this context is the level of spatial complexity in the sound design. Not all 30 positions in the above-mentioned example are relevant, and we aimed to identify a level of abstraction that would maintain a clarity of expression while still making use of this additional design dimension. Throughout this process, we experimented with various spatial effects that did not make it into the final design. The fundamental effect we noticed was that a sound's spatial movement path was influenced by the presence of a robot. Placing a robot in front of the listener led to sound in that general area be attributed to the robot. As a result, absolute sound positions, such as *left, middle, right* were more likely be perceived in relation to the robot (*outside the robot, inside the robot*). This manifested, among others, in the following ways:

- Making a sound circle around the listener is a popular technique in spatial sound design, because it highlights the spatial nature of the audio playback. When using this effect while a robot was in the field of view, sound positioned behind the robot was attributed to the robot. Any gradual circular movement was therefore broken up.
- More generally, the difference between sound located near the robot and sound located at its exact position was negligible. To tease out a difference in location between robot sound and environment sound, the distances had to be an estimated 3 meters or more.
- Rapid jumps between locations, such as the robot and an object in the environment, or simultaneous playback from multiple sources would perceptually fuse into a single global sound event and again be attributed to the robot, instead of being perceived as several distinct sources. This could be mitigated by either adding pauses between sound events at different locations, like a *call and response* pattern, or by giving the two sound sources significantly different timbres.

After excluding these and similar effects, we ended up working with a smaller number of key perceptually relevant locations and events. Those were a sound's *source*, *target*, and *transitions* between a source and target, or *across the entire space*. In other words, a sound was either *at* an object of interest, *moving away or towards* it, or it was *everywhere*. In the latter case, this could mean it was a global event fully in focus, or a subtle background ambient. These categories are reflected in a design framework in Section 5.2. Even with this rather reduced design space, the number of combinations is still substantial, allowing us to use spatial location to clearly differentiate sound events with similar timbres.

5.1.2 Applying Sound Installation Practice in Human-Robot Interaction

To conclude the lessons learnt, we will revisit the themes that emerged from the interviews in Section 3 to investigate how well they translated into actionable guidelines for the prototyping process and, by extension, how applicable and useful a sound installation perspective is to the HRI context.

When discussing **sonic identity and robot fiction**, experts emphasised how sound should be built around a core sound identity, noting how this helps create more coherent designs, while at the same time narrowing the design space. In the case of this prototype, we created a sound identity that took UX sound common among smart devices as a key reference. By mostly adhering to the conventions in that space, the sound set was relatively consistent. The key difference to more traditional UX sound was that we wanted to make a clear distinction between sound located in the robot, and sound emitted across the environment. We did so by making the sound inside the robot more mechanical and tactile, while making the sound in the environment more digital and smeared. An issue we faced in that context was that sound that was fitting for the sound identity was not necessarily fitting for the functions that sound needed to fulfil. An example of this was that many UX sound sets feature pure electronic tones with little high-frequency content. High-frequency content, however, was needed to make sounds easier to localise in space (see, for example, [13]). This presented a friction between functional and conceptual requirements for the design, and the audio processing and additional sound material we added to meet the functional requirements was one way to make this compromise. Generally, this theme's notions around identity, core fiction, and coherence were all reflected in the expert interviews in this chapter, prior robot sound expert interviews [33], our own prior design practice and the design work showcased in this chapter. We consider it highly applicable to the HRI context.

When discussing **functions of sound**, experts mentioned conveying emotion, providing orientation, and causing associations as key functions. They also suggested to choose sound materials based on the associations to be caused, and highlighted music and soundscapes as two unique ways to communicate beyond human speech. In the case of this prototype, we used musical elements throughout most of the design, taking traditional UX sound conventions and combining them with harmonic soundscapes. Providing orientation and drawing the listener's attention to specific locations was a key concern, and this was done through alert sounds with specific spatial properties. We also used sound material to create specific associations to a physical robot body (through tactile mechanical recorded sound) and to a disembodied digital entity (through heavily processed, smeared sound events). We consider these notions readily applicable to the HRI context. The possibility of using spatial sound to (i) subtly communicate through background ambient sound and (ii) guide the listener's attention to specific locations in the space, where both key motivations for the work in this chapter. While both of these have clear creative rationales – immersive musical soundscapes are pleasant to listen to, and space adds an exciting creative dimension to sound design – we also argue that those in particular have applications within more functional HRI contexts; that being able to subtly convey certain moods or draw attention to certain objects relevant to HRI scenarios is a valuable tool in an HRI designer's pallet.

When asked about **roles and affordances of interactive sound** – meaning sound that is in some way mapped to states and actions of human, robot, or the environment – the experts discussed various types of interactivity with varying degrees of complexity, illustrating their ideas with examples from their prior work. With this chapter's focus being on spatial sound, we implemented only basic interactivity. Key events were mapped to human presence. For example, approaching the robot body woke up the ambient presence and caused it to enter the physical robot. There were also several instances of mapping continuous robot movement data to the intensity of dedicated movement sounds. While we consider the various notions highlighted in the interviews to be generally applicable, the more complex ideas of interactivity were not explored in this prototype. Using distance as a parameter to accentuate human presence and robot awareness was, however, explored in previous work [34].

Discussing the **roles and affordances of distributed sound**, experts noted a significant increase in detail and complexity that comes with adding a spatial dimension to the sound design. They also highlighted its role in separating two sound sources to create noticeable differences between them even if they have similar timbres. One expert also suggested the use of spatial sound to separate an agent from its physical structure. In the case of this prototype we explored this complexity and found that when using it in the context of an agent, whether embodied or not, the space of potential sound locations had to be reduced, as not all design possibilities were equally relevant or effective. The use of spatial sound to create detailed immersive soundscapes that convincingly transport the listener to a different place is common in the installation context, but less relevant in this HRI context. The notion of separating sound sources was much more applicable. Clear distinction and disembodiment provided a wealth of design possibilities, and link well with HRI research into re-embodiment.

When asked about **technical considerations**, experts highlighted two key points, which partially contradicted each other: (i) To create quality designs, sound practitioners need access to a large number of loudspeakers to create a high-resolution spatial image. (ii) Rather than a complete high-resolution spatial image, designers only need the technological environment to create a few key experiences: enveloping sound or sound emitted from a point source, close or distant sound, and moving or static sound. In our prototyping work we chose to avoid potential technical constraints around spatial resolution by working within a virtual environment. Regarding the question of a high-resolution image versus a few key locations, the robot context provided interesting insights here. Due to the fact that sound was designed in the context of a robot, spatial locations were perceived in relation to it, which came with an inherent loss in spatial resolution. Some locations were inherently more relevant than others. Even though we had high-resolution spatial sound at our disposal, we ended up reducing the spatial resolution of our design, which leads us to believe that potential constraints in spatial resolution are less of an issue than some of the experts believe.

5.2 Design Framework

The core focus of this chapter – experimentation with a wide range of spatial behaviours – will now be generalised and simplified in the form of a design framework, a set of spatial sound distributions and trajectories we consider relevant to human-robot interactions. Its purpose is not to list or summarise the spatial sound work done for this specific prototype, but rather to map out a broader and simplified design space that explores the question "What are the ways distributed sound could be used in a robot's auditory communication?" These distributions and behaviours may then hopefully prove useful to HRI designers considering the use of spatial audio. The framework considers three *key locations* of robot sound (see Figure 8) which we consider as the most suitable simplification of the design space, and then applies this thinking to *four examples of spatial sound events* (see Figure 9). We argue that spatial robot sound should be broken down into these three key locations - the robot, objects of interest in the environment, and the space itself - and that transitions between these locations create spatial relationships which are relevant to HRI scenarios.

5.2.1 Key Locations

Based on the lessons learnt in Section 5.1.1 - not all sound locations are equally relevant to HRI and noticeable to the listener - we can render the broad space of possibilities down to three key sound-emitting locations. Figure 8 shows these locations: the robot, any position in the environment, and the space itself. If we imagine a social intelligence – let us again call it the robot mind – moving between these locations, we can also see them as states. The robot mind is either present within the robot body, at an arbitrary position in the environment, or everywhere. As a next step, we define transitions between these three states. While these are not strictly necessary, they allow the designer to convey continuity and direction between states. The language describing the transitions between these three states is heavily inspired by Smalley's spectromorphology [40]. While it ultimately references sound events, it uses descriptions of non-sounding phenomena.

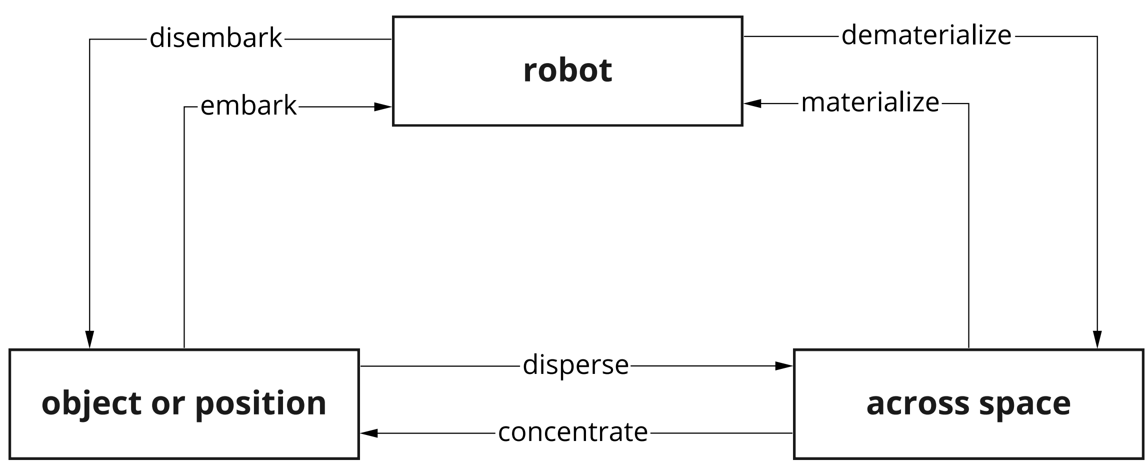

Figure 8: Three sound locations relevant to Human-Robot Interaction. Sound is emitted either by the robot itself, by an object or at any position in the environment, or across the space, without a discernible source.

Having the robot mind located within the **robot** body results in the standard scenario of an embodied agent. From there, the robot mind can *disembark* and travel to an arbitrary position or object in the environment. An example case for this would be the robot being asked to follow the user and continue a conversation in a room upstairs which the physical robot cannot access. To do so, the robot mind would disembark the robot body and move to a smart speaker in another room. The robot mind could also *dematerialize*, leaving the robot body to be present across the entire space. An example case of this would be the robot mind leaving the robot body to go into an inactive state, from which it can be awoken via voice commands.

Placing the robot mind at an arbitrary **object or position** brings the listening situation closer to interactions with current voice assistants in the home. The obvious object is, of course, a speaker, but voice interactions with other objects have been previously explored (see, for example, Iravantchi and colleagues' Digital Ventriloquism [23]). From this position, the robot mind can leave in two ways. It can leave its position and *embark* to become an embodied agent. An example case is a voice assistant taking on a physical form to do housework. The robot mind can also *disperse* to be present across the space, when it is not needed.

Lastly, the robot mind can be present **across the space**. In this case, there is no specific position the listener could discern as the sound source. Rather, sound is present in the entire space. From this state, the robot mind can *materialize* in the physical robot. An example case would be a robot mind on standby with a sonic presence across the space, which, after being activated by a voice command, confirms the command and subsequently moves into the robot body to carry out the request.

5.2.2 Four Examples of Spatial Sound Events

Keeping these key locations in mind, we now look at four examples for how a robot sound event could be designed spatially. To do so, we consider a sound moving between a source and a target. A source could either be a physical robot or a robot mind at any object or position in the environment. A target could be any object or location relevant the human-robot interaction scenario. Next to these two positions, we also consider the intensity of the sound used, as it allows us to emphasize and de-emphasize different locations along the sound's spatial trajectory. With these basic building blocks we can then attempt to convey relationships between the source and the target. Figure 9 shows a graphical representation of this. The simple visual language is based on the work of Blackburn [7], who created a rich and detailed visual representation of Smalley's spectromorphology, which we discussed in previous work [34]. It should be noted that the use cases in these examples are meant to illustrate the possible role of key sound locations, and not be an exhaustive list of possible applications.

Announce movement towards target - In this example, a robot plans a motion trajectory and announces it to any nearby listeners. The sound event begins with an alert sound located at the robot to draw attention to it, and then moves a second sound along the path the robot intends to take. Along this path, the sound loses intensity, placing emphasis on the initial robot location, while also indicating the general direction to robot plans to move in.

Convey pull from target - This example is an inversion of the above, but instead of placing emphasis on the initial robot location, emphasis is put on the goal and the message to be conveyed is not that the robot intends to move away from the source position, but instead that the robot is being pulled from a target

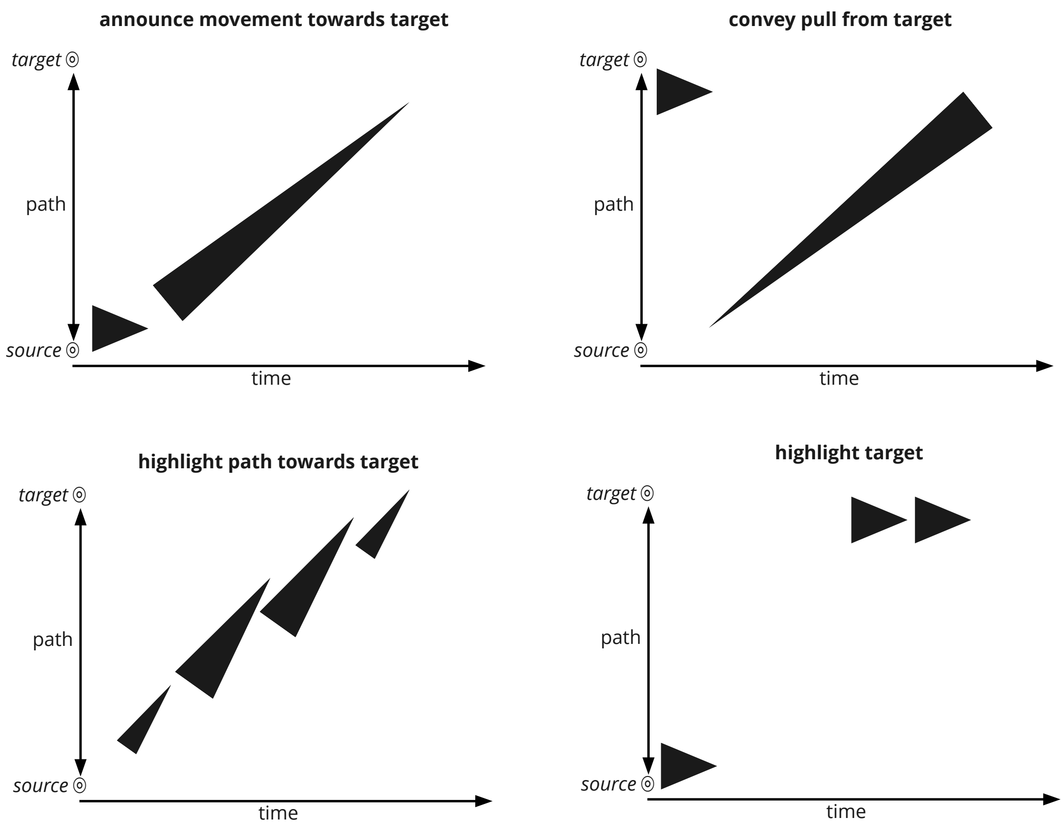

Figure 9: Four spatial sound events. Sound moves between source and target (y axis) over a certain amount of time (x axis) and changes intensity over time (thickness of shape).

position. To convey this, the listener's attention is first drawn to the target position, after which a second sound moves from the robot's initial location to the target while gradually increasing in intensity.

Highlight path towards target - In this example, we again have the situation of a robot announcing a movement path, but this time the actual path is the key information we want to bring across. Like an auditory equivalent of led strips in a plane that lead to the exit doors, we want to communicate the path a robot is going to take before it sets out on its journey. To do so, we move a sound from the source location to the target location and increase the sound's intensity along the path travelled. This then emphasises neither the current robot position, nor the target, but instead all the positions between the two.

Highlight target - Looking at events that do not feature sound movement, but simply timed sound events at different locations, we can consider the example of highlighting a target. In a situation where the robot wants to draw the human's attention to a certain object or position in the environment it could emit a sound at that position. We may also want to ensure that the human understands that this highlighting is part of their interaction with the robot. To do so, we could use a call and response model, where we first emit a sound at the robot position and then place a complementary sound event at the relevant position in the environment. By skipping the first step, we would still draw attention to the target, but it would be unclear who wanted to draw attention to it.

5.3 Limitations and Future Work

When considering the contributions of this chapter, some limitations should be noted, which are mostly based on the fact that the technology infrastructure for realising the experiences explored in this chapter at scale does not yet exist. There are currently no connected, distributed audio devices around robots, and therefore, a key underlying question is, how many of the affordances assumed by this work can be expected to manifest.

One possible limitation regarding the spatial resolution was previously mentioned in Section 5.1.2. While it is reasonable to be sceptical about the spatial resolution which might eventually be available to creators who design for this context, the core set of spatial experiences proposed in this chapter are still possible to achieve even when emitted across a loudspeaker system with reduced spatial resolution, and the design framework presented takes this into account. Another, potentially more severe limitation is the difference of timbre and balance across the devices making up the distributed audio system. While this challenge of unbalanced devices is primarily technical in nature, there are some design aspects to consider. In current work around Media Device Orchestration, the design of entertainment experiences takes the nature of their playback devices into account. This means that creators are encouraged to assign sound that requires full-range playback, like music, is sent to a device with the appropriate frequency range like a soundbar, while sound with a more limited spectrum, like a recording of a person whispering, can be sent across smaller devices like mobile phones. This practice is, in fact, reflected in this chapter's prototype, which assumes a full-range speaker with adequate low-end for the robot body itself, while all sound in the environment is designed to be thinner and can therefore be emitted by smaller speakers. However, this was evaluated in a virtual environment without any frequency range constraints and it therefore remains to be seen, how much this potential inconsistency across devices needs to be considered during the sound design process. In a worst case scenario one might imagine that individual audio devices in the home have such different timbral characteristics, that a sound moving from one corner of the room to another sounds so different throughout its journey, that the listener hears it as several devices each emitting their own, individual audio alert. This focus on individual devices as opposed to individual sounds moving through the space would make much of the effects explored in this chapter impossible to achieve.

While general impressions of the spatial work for this prototype will be gathered upcoming studies, there are a number of open questions to be addressed in future studies that could isolate and investigate specific effects. Relevant questions include (i) how different a robot-internal and robot-external sound set can be, before they are assigned to two separate agents, (ii) how close a sound can be to a physical robot, before it is attributed to it, and (iii) how much sound design complexity people consider appropriate in a real-world context with other people and sound sources present.

6 Conclusion

In this chapter, we explored the role of spatially distributed sound in human-robot interaction by creating a spatial sound set to be emitted by a robot and its environment. We first interviewed researchers and practitioners who work with distributed audio and sound installations and identified five themes relevant to distributed sound in the context of HRI: *sonic identity and fiction*, *functions of sound*, *roles and affordances of interactive sound*, *roles and affordances of distributed sound*, and *technical considerations.* We then combined these themes with our own experience in interactive immersive audio environments to create a distributed robot sound prototype. The prototyping process was comprised of five stages: defining *fiction*, developing a *sound identity*, creating an *audio interaction flow*, *asset production*, and *implementation* into a virtual simulation environment. We then reflected on lessons learnt during this design process, evaluating the various tools and approaches we used. Finally, we proposed a generalised design framework for spatial robot sound, arguing that spatial robot sound can be broken down into three key locations - the robot, objects of interest in the environment, and the space itself - and that transitions between these locations create spatial relationships which are relevant to HRI scenarios.

References

[1] Albastaki, A., Hoggenmueller, M., Robinson, F. A., and Hespanhol, L. Augmenting Remote Interviews through Virtual Experience Prototypes. In *32nd Australian Conference on Human-Computer Interaction (OzCHI '20), Sydney, NSW, Australia* (New York, NY, USA, Dec. 2020), ACM, p. 9. ZSCC: 0000001.

[2] Alsaawi, A. A critical review of qualitative interviews. *European Journal of Business and Social Sciences 3*, 4 (2014). ZSCC: 0000214.

[3] Atzori, L., Iera, A., and Morabito, G. The Internet of Things: A survey. *Computer Networks 54*, 15 (Oct. 2010), 2787–2805. ZSCC: 0012847.

[4] BBC. Audio Orchestrator, Sept. 2021.

[5] Bechhofer, F., and Paterson, L. *Principles of research design in the social sciences.* Routledge, 2012. ZSCC: 0000544.

[6] Berg, B. L., Lune, H., and Lune, H. Qualitative research methods for the social sciences. ZSCC: 0032155 Publisher: Pearson Boston, MA.

[7] Blackburn, M. The Visual Sound-Shapes of Spectromorphology: an illustrative guide to composition. *Organised Sound 16*, 1 (2011), 5–13. ZSCC: 0000063 Publisher: Cambridge University Press.

[8] Bordignon, M., Rashid, J., Broxvall, M., and Saffiotti, A. Seamless integration of robots and tiny embedded devices in a PEIS-Ecology. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems* (San Diego, CA, USA, Oct. 2007), IEEE, pp. 3101–3106.

[9] Braun, V., and Clarke, V. Using thematic analysis in psychology. *Qualitative Research in Psychology 3*, 2 (Jan. 2006), 77–101. ZSCC: 0071387.

[10] Broxvall, M., Gritti, M., Saffiotti, A., Beom-Su Seo, and Young-Jo Cho. PEIS Ecology: integrating robots into smart environments. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.* (Orlando, FL, USA, 2006), IEEE, pp. 212–218.

[11] Bui, H.-D., and Chong, N. Y. An Integrated Approach to Human-Robot-Smart Environment Interaction Interface for Ambient Assisted Living. In *2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)* (Genova, Italy, Sept. 2018), IEEE, pp. 32–37. ZSCC: 0000004.

[12] Cain, R., Jennings, P., and Poxon, J. The development and application of the emotional dimensions of a soundscape. *Applied Acoustics 74*, 2 (Feb. 2013), 232–239. ZSCC: 0000162.

[13] Carlile, S., Delaney, S., and Corderoy, A. The localisation of spectrally restricted sounds by human listeners. *Hearing research 128*, 1-2 (1999), 175–189. ZSCC: 0000071 Publisher: Elsevier.

[14] Cha, E., Fitter, N. T., Kim, Y., Fong, T., and Matarić, M. J. 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* (Chicago, IL, USA, 2018), ACM Press, pp. 434–442. ZSCC: 0000006.

[15] De Marco, R. *Sound Scenography: The Art of Designing Sound for Spaces.* Avedition Gmbh, Stuttgart, Oct. 2021. ZSCC: NoCitationData[s0].

[16] Do, H. M., Pham, M., Sheng, W., Yang, D., and Liu, M. RiSH: A robot-integrated smart home for elderly care. *Robotics and Autonomous Systems 101* (Mar. 2018), 74–92.

[17] Drury, J., Scholtz, J., and Yanco, H. Awareness in human-robot interactions. In *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)* (Washington, DC, USA, 2003), vol. 1, IEEE, pp. 912–918. ZSCC: 0000307.

[18] Francombe, J., and Hentschel, K. Evaluation of an immersive audio experience using questionnaire and interaction data. In *Proceedings of the 23rd International Congress on Acoustics* (Aachen, Germany, 2019), p. 8. ZSCC: 0000006.

[19] Francombe, J., Mason, R., Jackson, P. J., Brookes, T., Hughes, R., Woodcock, J., Franck, A., Melchior, F., and Pike, C. Media Device Orchestration for Immersive Spatial Audio Reproduction. In *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences - AM '17* (London, United Kingdom, 2017), ACM Press, pp. 1–5. ZSCC: 0000008.

[20] Francombe, J., Woodcock, J., Hughes, R., Mason, R., Franck, A., Pike, C., Brookes, T., Davies, W., Jackson, P., Cox, T., Fazi, F., and Hilton, A. Qualitative Evaluation of Media Device Orchestration for Immersive Spatial Audio Reproduction. *Journal of the Audio Engineering Society 66*, 6 (June 2018), 414–429. ZSCC: NoCitationData[s1].

[21] Francombe, J., Woodcock, J., Hughes, R. J., Hentschel, K., Whitmore, E., and Churnside, T. Producing Audio Drama Content for an Array of Orchestrated Personal Devices. In *Audio Engineering Society Convention 145* (2018), Audio Engineering Society. ZSCC: 0000005.

[22] Gomm, R. *Social research methodology: A critical introduction.* Macmillan International Higher Education, 2008. ZSCC: 0001093.

[23] Iravantchi, Y., Goel, M., and Harrison, C. Digital Ventriloquism: Giving Voice to Everyday Objects. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), pp. 1–10. ZSCC: 0000004.

[24] Kiesler, S., Powers, A., Fussell, S. R., and Torrey, C. Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition 26*, 2 (2008), 169–181. ZSCC: 0000280 Publisher: Guilford Press.

[25] Kim, Y. J., Kim, B. H., and Kim, Y. K. Surround audio device and method of providing multichannel surround audio signal to a plurality of electronic devices including a speaker, Dec. 2019. ZSCC: 0000000.

[26] Lawless, B., and Chen, Y.-W. Developing a method of critical thematic analysis for qualitative communication inquiry. *Howard Journal of Communications 30*, 1 (2019), 92–106. ZSCC: 0000078 Publisher: Taylor & Francis.

[27] Leavy, P. *The Oxford handbook of qualitative research.* Oxford University Press, USA, 2014. ZSCC: 0000634.

[28] Li, J. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies 77* (2015), 23–37. ZSCC: 0000279 Publisher: Elsevier.

[29] Loupa, G. Influence of Noise on Patient Recovery. *Current Pollution Reports* (2020), 1–7. ZSCC: 0000000.

[30] Luria, M., Reig, S., Tan, X. Z., Steinfeld, A., Forlizzi, J., and Zimmerman, J. Re-Embodiment and Co-Embodiment: Exploration of social presence for robots and conversational agents. In *Proceedings of the 2019 on Designing Interactive Systems Conference - DIS '19* (San Diego, CA, USA, 2019), ACM Press, pp. 633–644.

[31] Mason, R. How Important Is Accurate Localization in Reproduced Sound? In *Audio Engineering Society Convention 142* (2017), Audio Engineering Society.

[32] Reig, S., Luria, M., Forberger, E., Won, I., Steinfeld, A., Forlizzi, J., and Zimmerman, J. Social Robots in Service Contexts: Exploring the Rewards and Risks of Personalization and Re-embodiment. In *Designing Interactive Systems Conference 2021* (2021), pp. 1390–1402. ZSCC: 0000000.

[33] Robinson, F. A., Bown, O., and Velonaki, M. Designing Sound for Social Robots: Candidate Design Principles. *International Journal of Social Robotics* (June 2022).

[34] Robinson, F. A., Velonaki, M., and Bown, O. Crafting the Language of Robotic Agents: A vision for electroacoustic music in human–robot interaction. *Organised Sound* (2022), 1–13. Publisher: Cambridge University Press.

[35] Roginska, A., and Geluso, P., Eds. *Immersive sound: the art and science of binaural and multichannel audio.* Routledge, Taylor & Francis Group, New York ; London, 2018. ZSCC: NoCitationData[s0].

[36] Rumsey, F., Zieliński, S., Kassier, R., and Bech, S. On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *The Journal of the Acoustical Society of America 118*, 2 (Aug. 2005), 968–976.

[37] Sayin, E., Krishna, A., Ardelet, C., Briand Decré, G., and Goudey, A. "Sound and safe": The effect of ambient sound on the perceived safety of public spaces. *International Journal of Research in Marketing 32*, 4 (Dec. 2015), 343–353. ZSCC: 0000028.

[38] Schafer, R. M. *The new soundscape.* BMI Canada Limited Don Mills, 1969. ZSCC: 0000602.

[39] Seidman, I. *Interviewing as qualitative research: A guide for researchers in education and the social sciences.* Teachers college press, 2006. ZSCC: 0000029.

[40] Smalley, D. Spectromorphology: explaining sound-shapes. *Organised sound 2*, 2 (1997), 107–126. ZSCC: 0000945.

[41] Turcu, C., Turcu, C., and Gaitan, V. Integrating robots into the Internet of Things. *International Journal of Circuits, Systems and Signal Processing 6*, 6 (2012), 430–437.

[42] Wonpil Yu, Jae-Yeong Lee, Young-Guk Ha, Minsu Jang, Joo-Chan Sohn, Yong-Moo Kwon, and Hyo-Sung Ahn. Design and Implementation of a Ubiquitous Robotic Space. *IEEE Transactions on Automation Science and Engineering 6*, 4 (Oct. 2009), 633–640. ZSCC: 0000035.

[43] Woodcock, J., Francombe, J., Hughes, R., Mason, R., Davies, W. J., and Cox, T. J. A quantitative evaluation of media device orchestration for immersive spatial audio reproduction. In *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science* (2018), Audio Engineering Society. ZSCC: 0000010.