A Taxonomy of Sounds in Virtual Reality

Dhruv Jain
Microsoft Research, Redmond, Washington; University of Washington, Seattle, Washington
djain@uw.edu

Sasa Junuzovic
Microsoft Research, Redmond, Washington
sasajun@microsoft.com

Eyal Ofek
Microsoft Research, Redmond, Washington
eyalofek@microsoft.com

Mike Sinclair
Microsoft Research, Redmond, Washington
sinclair@microsoft.com

John R. Porter
Microsoft Research, Redmond, Washington
joporte@microsoft.com

Chris Yoon
Microsoft Research, Redmond, Washington
chyoon@microsoft.com

Swetha Machanavajhala
Microsoft Research, Redmond, Washington
swmachan@microsoft.com

Meredith Ringel Morris
Microsoft Research, Redmond, Washington
merrie@microsoft.com

ABSTRACT

Virtual reality (VR) leverages human sight, hearing and touch senses to convey virtual experiences. For d/Deaf and hard of hearing (DHH) people, information conveyed through sound may not be accessible. To help with future design of accessible VR sound representations for DHH users, this paper contributes a consistent language and structure for representing sounds in VR. Using two studies, we report on the design and evaluation of a novel taxonomy for VR sounds. Study 1 included interviews with 10 VR sound designers to develop our taxonomy along two dimensions: sound source and intent. To evaluate this taxonomy, we conducted another study (Study 2) where eight HCI researchers used our taxonomy to document sounds in 33 VR apps. We found that our taxonomy was able to successfully categorize nearly all sounds (265/267) in these apps. We also uncovered additional insights for designing accessible visual and haptic-based sound substitutes for DHH users.

CCS CONCEPTS

• **Human-centered computing** → Accessibility.

KEYWORDS

accessibility, virtual reality, deaf, Deaf, hard of hearing, taxonomy, sound awareness, audio engineering, sound design

ACM Reference Format:
Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John R. Porter, Chris Yoon, Swetha Machanavajhala, and Meredith Ringel Morris. 2021. A Taxonomy of Sounds in Virtual Reality. In *Designing Interactive Systems Conference 2021 (DIS '21), June 28–July 02, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3461778.3462106

1 INTRODUCTION

Virtual reality (VR) environments are filled with a rich diversity of sounds ranging from sounds that provide critical notifications (*e.g.,* video-game "enemy" footsteps) to those that increase realism (*e.g.,* wind blowing in a nature simulation) [2, 6]. For many d/Deaf and hard of hearing (DHH) people[1], these sounds, and the information they convey may not be accessible, which may limit their VR experience (*e.g.,* by missing critical cues in games or conversations in social apps). Our research explores how to make sounds in VR accessible to DHH people. This goal is multifaceted, and to begin with, requires a thorough understanding of how different sounds are used and represented in VR. In this paper, we present a design and evaluation of a novel taxonomy to organize and discuss VR sounds with a long-term goal to make them accessible to DHH people.

While prior work has articulated several taxonomies of sounds, the focus has largely been on organizing real-life sounds such as animal calls [9, 35], human speech [4, 32], and music [18, 37]. These taxonomies are not amenable for VR worlds, which also include many synthetic sounds (*e.g.,* Foley [52]) and exaggerated representations of real-life sounds [7]. A few sound taxonomies exist for (2D) video games and films [14, 15, 47], which relate more closely to VR than real-world sound taxonomies. However, they have several limitations. First, these taxonomies either cover the source (*e.g.,* a character speaking) or the intent (*e.g.,* to convey critical information, to increase realism) of sounds, but not both [15, 19]. Accounting for only one dimension often leads to ambiguity in certain nuanced meanings of sounds—for example, the intent of a character's speech (sound source) could be to convey critical cues for progression in a game or could just be street noise for increasing realism. Second, prior taxonomies were constructed for two dimensional screens and, in contrast to our taxonomy, fail to account for 3D spatial variations in VR (*e.g.,* a character speaking in

DIS '21, June 28–July 02, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8476-6/21/06…$15.00
https://doi.org/10.1145/3461778.3462106

---

[1] Two of our paper authors, including the first author, are DHH.

front of a player vs. a background narrative speech). Finally, prior taxonomies were built for specific applications (*e.g.,* action games [47] or animation movies [14]) while our taxonomy is broadly applicable across different VR app categories.

To design and evaluate our taxonomy, we performed two studies. First, we interviewed 10 VR sound designers who described their experience of designing sounds for VR apps including the different types and characteristics of VR sounds. Through qualitative analysis, we developed our novel sound taxonomy consisting of two dimensions: sound source (*e.g.,* inanimate objects, point ambience) and intent (*e.g.,* for conveying critical information, for inducing affective state) as well as uncovered additional themes that may influence VR sound accessibility for DHH users (*e.g.,* characteristics such as volume, persistence, and spatial location as well as whether the sound is accompanied by visual or haptic feedback). To evaluate our taxonomy across different VR apps, we conducted a second study where eight hearing HCI researchers used 3-5 VR apps each (a total of 33 apps across 10 categories) and classified the sounds in these apps using our proposed taxonomy. We found that our taxonomy was able to describe nearly all sounds (265/267) in these apps. We also identified what sound categories are important to represent accessibly and design considerations for developing accessible visual and haptic augmentations for these sounds.

In summary, our work contributes: (1) a novel taxonomy that articulates both the source and intent of sounds across different VR apps, and (2) additional insights from two studies with 10 sound designers and 8 HCI researchers including several design considerations for making VR accessible to DHH users.

2 BACKGROUND AND RELATED WORK

We provide background on DHH culture as well as discuss prior work in sound taxonomies and VR accessibility.

2.1 Background on DHH Culture

For many DHH people, the degree of hearing loss is only a small aspect of their disability and does not determine their choice of language or accessible technologies [6, 27, 31]. To understand what factors affect inclusion, researchers have composed three models of deafness: medical, social and cultural [6, 54]. In the medical model, a person with hearing loss is seen as wanting to restore normal hearing. In the social model, a DHH individual is considered as needing to integrate into the society of hearing people. Finally, in the cultural model, a DHH person is viewed as part of a culture or community with a distinct visual language. Usage of these models depends on the research goals [6, 27]. For example, to develop hearing aids and cochlear implants for (partially) restoring hearing, researchers primarily embody the medical model [6]. Here, we leverage the social and cultural perspectives of DHH people when addressing VR accessibility.

Within the cultural model, an individual can identify as deaf, Deaf (capital 'D'), or hard of hearing. The term "Deaf" refers to people who belong to a Deaf culture with common language, values and practices (see [6, 28, 31] for details). In contrast, the terms "deaf" and "hard of hearing" indicate someone for whom deafness is primarily an audiological experience and who refrain from membership to a particular community [6, 31]. Individuals belonging to these groups do not have a distinct cultural identity of their own, and they may choose to interact with either hearing or Deaf people based on their comfort [6, 31]. Our work addresses all three DHH groups.

The preferences for sound information may differ across DHH cultural groups. For example, prior large-scale surveys for real-life sounds [2, 13] have shown that people who prefer oral communication may be more interested in some sounds (*e.g.,* phone ring, conversations) than those who prefer sign language. Nevertheless, both groups have uniform preferences for characteristics of sounds, desiring some (*e.g.,* identity, location) more than others (*e.g.,* loudness, pitch), and both prefer using both visual and vibration modalities for receiving sound feedback [2, 13]. These preferences may also influence the design of VR sound accessibility.

2.2 Sound Taxonomies

To assist with the growing sound classification research, past work—including specifically for DHH people [2, 13, 23]—has largely focused on identifying taxonomies of real-life sounds for different contexts and use cases, such as restaurants [29], urban areas [16, 43], animals [9, 35], speech [4, 32], and music [18, 37]. We instead detail taxonomies of sounds in video games and films, which relate more closely to VR than real-world sounds.

Three decades ago, a common classification of (2D) video game sounds included three categories: speech, sound, and music [3, 19]. This taxonomy was limited and, for example, could not distinguish between background and foreground sounds [19] such as music in the background *vs.* music from in-game elements such as a piano. Folmann [14] extended this classification by distinguishing background and foreground sounds (a common categorization of film music [8, 50]), leading to four categories: voice, sound effects (foreground), ambience (background), and music. Although valuable, this taxonomy only reveals the source of the sound and not the intent for the player's experience [15, 19]. Similarly, Stockburger [47]—who introduced the concept of diegesis to distinguish sounds that are linked to objects in the game world (*e.g.,* footsteps, car engines) from those that are not (*e.g.,* background music, interface clicks)—divided the game audio based on "sound objects": speech (dialogue), zone (ambience), score (music), effect (diegetic game sounds), and interface (non-diegetic menu sounds). While employed for nearly a decade in audio production process [19, 26], this classification also does not easily reveal the sound intent.

To reveal the intent of the sounds, Friberg and Gärdenfors [15] proposed a categorization based on organization of sound assets in the game—that is, avatar sounds, object sounds, character sounds, ornamental sounds, and instructions. However, this approach applied only to specific avatar-based games, and seldom disclosed the source of the sound [19, 47]. Several iterations of Friberg and Gärdenfors' taxonomy exists (*e.g.,* IEZA [19]), although each with its notable criticisms (*e.g.,* [1, 26]).

In summary, the prior taxonomies either reveal the source or the intent of sounds, but not both, which—as described in the Introduction—is essential to accurately represent different sound categories. Moreover, these taxonomies were built for 2D films and games and have not considered virtual reality environments, which includes other app categories beyond games (*e.g.,* social apps, educational content, simulated experiences) and 3-D interactive audio that reacts to the user's position and actions instead of the "locked-in" track in films and 2-D games [57]. Using interviews with VR sound designers and an inventory of sound use in VR apps, we investigate a sound taxonomy that is specifically tailored to VR, is useful for DHH accessibility, and accounts for both the origin of sound and its intended functionality.

2.3 VR Accessibility

Explorations of accessibility in VR have only recently gained traction. For example, the first two symposiums for virtual, augmented, and mixed reality accessibility (*XRAccess* [58]) were held in 2019 and 2020. Informed by the symposium discussions, the World Wide Web Consortium (W3C) published the first working draft of VR accessibility requirements in Feb 2020 [59]; relevant needs for DHH users include customizable subtitles, description of important sound events, and availability of binaural audio recordings.

Prior efforts in VR accessibility have largely addressed people with visual [46, 49, 55] or mobility impairments [17, 38]. For DHH users, researchers have explored the design of specialized VR apps, for example, to assist with storytelling [12, 48] or teaching [36, 40]. While not formally evaluated in the literature, some commercial VR games also offer accessibility features for DHH users such as subtitles [60] or sound source direction indicators [61]. By holistically characterizing sounds across VR apps, our work focuses on mainstream VR accessibility, which has advantages of lower development costs, increased availability, and better social acceptability over specialized VR experiences; the latter risk stigmatization and technology abandonment [39, 42, 45].

While not focused on VR, prior work has explored real-life sound visualizations for augmented reality (AR)headsets, such as for showing captions [20, 22, 41] or sound source direction [21]. While these efforts may inform VR accessibility, sounds in VR differ from real-life (due to presence of fictional and exaggerated real worlds) and require a separate investigation from the use of sound in AR.

3 STUDY 1: DESIGN OF A TAXONOMY FOR VR SOUNDS

To investigate how sounds are designed and used in VR, we conducted a semi-structured interview with 10 VR sound designers. By critically analyzing the responses, we articulated a taxonomy that categorizes VR sounds according to two dimensions: sound source and sound intent. Our choice of taxonomy generation was empirical—instead of, for example, using a design framework (*e.g.,* [5]) or analysis of prior work (*e.g.,* [33])—because VR sound design is still in its infancy and not much is known about VR sounds.

3.1 Method

Participants: We recruited 10 sound designers using professional email lists and snowball sampling (Table 1). Although we did not collect data on this, professional sound designers usually have degrees in sound engineering, digital media, or game art and design and have experience with audio editing and music production. All designers were U.S. residents, were hearing, and had worked on at least one publicly released VR project.

Table 1: Demographics of VR sound designers in Study 1.

ID	Age	Gender	VR sound design experience (in yrs.)
D1	49	male	5
D2	24	male	1.5
D3	26	female	4
D4	23	female	3
D5	29	female	5
D6	31	male	1
D7	43	male	6
D8	53	female	2
D9	33	male	2.5
D10	35	male	4

Procedure: Due to the COVID-19 pandemic, all interviews were conducted over videoconference by a hard of hearing team member[2]. To facilitate accessibility for the interviewer, a professional transcriptionist attended all sessions. The interviewer asked about the designers' roles and responsibilities for up to three VR projects, including the different sounds used (*e.g.,* object sounds, people sounds), different aspects of the sounds considered for the design (*e.g.,* spatial location, environmental effects), differences between VR and non-VR sound design, accessibility considerations used when designing sounds for VR apps, and audio design rules or languages followed. The interviews lasted up to one hour and participants were compensated $25.

Data analysis: We retained the professional interview transcripts and used an open, axial, and selective coding process [10] to generate the taxonomy and other relevant themes (*e.g.,* sound characteristics, sound design guidelines). One team member read the transcripts and created open codes to summarize the data, resulting in 17 codes for the taxonomy (*e.g.,* speech, objects sound) and 22 codes for other themes (*e.g.,* audio file formats). The team member then split, merged, and reorganized the open codes to form meaningful relationships and to generate mutually exclusive taxonomy categories. For example, the "speech" category was split into "localized speech" and "non-localized speech" to incorporate the possible 3D spatial variation in speech (*e.g.,* a character speaking *vs.* ambient speech). Similarly, sounds from "non-speech human activities" and "animal activities" were combined into "animate objects" because of potential overlap between the two categories (*e.g.,* footsteps guidelines). This process generated 14 axial codes for the taxonomy (*e.g.,* localized speech) and 11 other axial codes (*e.g.,* technical aspects of audio). Finally, the axial codes were combined into overarching themes, resulting in two dimensions for the taxonomy (sound source and sound intent) and five other parent themes (*e.g.,* sound design guidelines, accessibility of VR apps). This taxonomy, themes, and containing codes were then shared with the other research team members for discussion and refinement and example excerpts for each code were collected from the transcript. Once finalized, the taxonomy was verified with the 10 sound designers over email.

---

[2] Both studies described in this paper were approved by our organization's IRB.

3.2 Findings

We describe the organization and characteristics of sounds in VR, differences between VR and real-life sound design, standard approaches to VR sound design, and accessibility of sounds in existing VR apps. These insights were compiled from 10 sound designers' self-reports of about 25 VR apps they had worked on spanning multiple categories such as art (*e.g.,* drawing tools), social (*e.g.,* a virtual meetup), travel (*e.g.,* a city scene), games (*e.g.,* shooting games), sports (*e.g.,* air hockey), and music (*e.g.,* virtual concerts). Quotes are drawn from the interview transcripts.

Organization of sounds in VR: We found that there is no agreed-upon terminology or common language for describing sounds in VR; consequently, each designer had to develop their own sound categorization based on preferences and project needs. D8 outlined this issue:

> “We were trying to describe the sound when a car engine sounded weird. [...] And we described it as a “chuddering” sound. It sounded like an onomatopoeia for it, but that is it. [...] No one else is going to know it. That's how a lot of audio stuff goes by in this field. Because people name different things. And they don't have a common language for it. [...] And it is like a big deal when someone comes up with a nice set of 12 terms to describe sounds in a concert hall. [...] This should be a solved problem. But it is still a subject of research.”

Nevertheless, through qualitative coding on the interview responses, we were able to arrive at a taxonomy of sounds that all 10 designers agreed with (when shown to them *post-hoc* over email). The two dimensions of the taxonomy appear in Table 3 and are described below. Similar to Stockburger [47], we use the term *diegetic* sounds to refer to the sounds emanating from objects in the game world (*e.g.,* a river flowing) and non-diegetic sounds for sounds that are played in the background (*e.g.,* a background music).

Sound source: This dimension identifies the (virtual) source that produced sound. The 9 categories include:

- *Localized speech:* spatially positioned speech (*e.g.,* a character speaking)
- *Non-localized speech:* ambient speech (*e.g.,* a narrator, player thinking aloud).
- *Inanimate objects:* sounds from non-living objects (*e.g.,* weapons, appliances)
- *Animate objects:* non-speech sounds from living beings (*e.g.,* footsteps, animal calls)
- *Interaction sounds:* sounds from interaction between multiple objects (*e.g.,* player touching a menu or punching an enemy)
- *Point ambience:* spatialized ambient sounds that are diegetic—that is, belong to the game world (*e.g.,* a river on one side of the player)
- *Surrounding ambience:* diegetic non-spatialized ambient sounds (*e.g.,* a crowd)
- *Notification sounds:* non-diegetic critical alerts (*e.g.,* low on ammunition, end of a player's turn).
- *Music:* non-diegetic background music

Note that these categories are mutually exclusive—for example, “interaction sounds” (*e.g.,* player touching a menu) and “inanimate objects” (*e.g.,* weapons) may appear to overlap with “notification sounds” (*e.g.,* end of a player's turn, low on ammunition), but the former two categories represent diegetic sounds [47] while the latter are non-diegetic background sounds.

Sound intent: This dimension describes the intended functionality of the sound. We detail the five categories below; the sixth category in Table 3 was added after Study 2.

- *Sounds for conveying critical information:* all sounds that are important for progression in an app (*e.g.,* enemy footsteps, low on ammunition, end of a player's turn). For example,

"So the sounds of those mobs, the zombies and skeletons, are very critical. Because you don't want to get killed. So you have to be able to hear the skeleton or zombie coming to you, and when they are shooting at you. [. . .] And lava sounds are critical. Because of course you get burnt when you go into the lava." (D9)

- *Sounds for increasing realism:* ambient or objects sounds that increase immersion (*e.g.,* river flowing, vehicles). For example,

“if it's [a] city, you will have a car and like traffic ambience sounds. And if it's like day or if it's nighttime, the sound will be different too. Day time has a crowd [...] people on [the] streets, while nighttime is quiet.” (D10)

- *Sounds for rhythm or movement:* music in an exercise or dancing app. For example,

"when doing aerobics, there are music beats like thic, thic, thic, thic. . ." (D3)

- *Sounds for generating an affective state*: emotional sounds (*e.g.,* stressful sounds when approaching the end of a level, calm music in a meditation app). For example,

“The music evolves depending on the intensity of the game. So, level 1 will just be kind of ambient synth pads, just very slow string or bass build up. There is no beat to it because nothing is too scary yet. [...] And as you get more intense, there are percussion sounds. Boom, boom, boom. And more instructions to build up that fullness. But not too fast. And as you get intense, there's like an electric rock guitar coming in. And more beats and will be faster too. So, music evolves with the emotion the player will be feeling.” (D4)

- *Sounds for aesthetics or decoration:* non-critical sounds that increase beauty (*e.g.,* background music, sounds accompanying decorative visuals). For example,

“[In an art app,] we had an eraser. To make it cartoony and playful and friendly, we added little effects: [...] kind of like pink puffs coming off from behind it. And we used audio to give the puffs a sense of tone.” (D3)

Sound Characteristics: Besides the sound categories, designers also reported on characteristics of sounds they considered while designing sounds in VR such as identity (*e.g.,* footsteps, gunshot), loudness, pitch, persistence, spatial location, priority, and environmental effects (*e.g.,* reverberation, spatial blend). We explain the three less obvious terms (persistence, spatial location, and priority) below.

Persistence refers to how long the sound is active. While most sounds have realistic durations (*e.g.,* a dog bark sound in VR will resemble the duration of real dog's barking), designers sometimes increase persistence to gain the user's attention. For example,

> "Sometimes in VR, you have something behind you and you need to tell the person to look over there. That's when sound design sort of turns into like attention design and it needs to be long enough of a sound that they have time to not only attend to it but interact with it or find it within this geography, right? [...] For things that are in front of you, they can be [of] short [duration], like a normal real [sound]. But if it needs to be indicating something [important], you need to play with duration or pitch [...] to indicate that something needs your attention. So we use persistent audio as opposed to transient audio. Sometimes it is like a drone, or sometimes it is a loop. So like in [a navigation app for blind users], the [directional] beacon we use is a 3-second long asset that we just loop, loop, loop. [...]" (D7)

Spatial location includes the sound's point location and the spread. For example,

> "For the gondola app, the sound was coming from up there to you. So, I assigned an attenuation curve from the position of the sound emitter, the gondola, to the listener. How much the sound decreased in the space. [. . .] Plus there was wind and all so you will hear less of it than in if it were in vacuum. . ." (D7)

Finally, *priority* conveys the importance of a sound. In case of multiple simultaneous sounds, priority helps filter low-priority sounds to prevent overwhelming the user or overloading the hardware resources. For example,

Let's say you [are in] a [sandbox game] world and you have like 100 sheep in front of you. [...] You have to prioritize what sounds are more important than other sounds. You don't want to overload the CPU or overwhelm whoever is playing. So, for example, the hostile mob sounds, you want to be able to hear it more clearly than like the sheep's footsteps sound, which are not really critical. [...] So I would make the attenuation curve a little more aggressively on the sheep's footstep sounds, so you don't really hear it. Unless you are really close to the sheep..." (D4)

VR sounds *vs.* real-life sounds: To assess if we can apply design principles from prior work in real-life sound feedback to VR, we also asked about similarities and differences between VR and real-life sounds. The designers mentioned many differences between VR and real-life sound design due to user experience adjustments, hardware limitations, and because VR enables artificial experiences. For user experience, D3 explained that since VR is a new technology, users may not be able to distinguish sounds as well as in real-life:

> "[During] the meetings in AltSpace [...], if somebody's [speech] volume is low, people can't tell if they are talking softly or are they far away? Whereas in the real world [...], [they] can instantly tell if you are close to [them] and are talking quietly, or you are far away. [. . .] Our brains are good at prioritizing [...] [and] saying oh, well, this thing is closer to me so I will listen to this first. But that is hard to do in VR. I guess it is so new and people are still getting to know it. And then the speakers have all sorts of problems [...] So it becomes really hard to do meetings if you have 100 avatars around you and everybody is talking at the same time..."

Due to this limitation, designers make creative decisions for sounds in VR to be more pleasant and manageable. For example, related to the above quote, D3 added:

> "So, we have a feature called a megaphone. And that says no matter how far you and I are, if you're [looking at] me, you hear me [. . .] at 100% volume. But if I am looking elsewhere but still in your visual range, then volume is maybe 20% less. [...] And if I am [behind you], but still near, maybe the volume you hear is 50% less."

Another related adjustment is to convey only a few selected sounds in a busy scene to reduce cognitive overload. For example,

> "We have to be super careful about like oh, you know when the user is shooting, then everything else will have to be reduced by a certain decibel to keep the experience clean. [...] It is still chaotic, but you want them to be able to focus, like [if] it is only these two that are fighting, then everything else will have to be quieter." (D2)

Finally, to keep the experience more engaging, sounds are sometimes designed to be more aesthetically pleasing than in real-life. D8 explains this:

> "A lot of rooms sound boring [in real-life]. So, part of what the sound designer is sometimes doing is creating a sound model that isn't strictly related to the real physics [...] to create a sound that is partially grounded in reality but which sounds more compelling. You [. . .] take the physical model and sort of modify it artistically."

Besides user experience, hardware limitations such as limited frequency range, limited audio channels, and less computing power (CPU/GPU) also force designers to make adjustments that deviate from real-life sounds:

> "You're playing a mobile [VR] game, you are not going to have a very nice headphone, like HTC Vive so you kind of have to make sure that the frequency range works for your crappy little speaker on your mobile phone. Or the earbuds or AirPods. So, I wouldn't use very low frequency heavy sounds. [...] And I wouldn't like too sharp of a sound because mobile speakers already make sounds a little sharp-y. So, I would cut high frequencies." (D4)

Beyond depicting reality, VR can enable artificial experiences, and in the absence of other modalities (*e.g.,* touch, smell) adding some creative Foley sounds [52] helps make those experiences immersive. For example,

> "If in VR I tap, and a panel pops up. There are no real-life [objects] that I can relate to [for] tapping. [...] You are tapping air... So, any time when you are tapping air, you do want the user to feel the satisfaction of hitting a button and it is kind of hard to achieve. [...]" (D6)

Another example relates to unrealistic motions in VR, which requires unrealistic sound designs:

> "For VR, because you have the controllers in your hand, you can teleport and like jump — so if you are implementing music, you would have to interpolate things differently than you would in AR [augmented reality] where you don't expect people to hop. [...] You want [the sounds before/after a jump in VR] to feel more naturalistic, like with the sky view thing where we were [...] in Sweden trying to tell these really naturalistic stories and we were hoping that you will almost forget that you had headphones on because the idea was like you are here, looking at the gondola at a distance and we want to kind of almost transport you naturalistically inside that [gondola] for 30 seconds and then teleport you back and then you can listen to another story." (D2)

This example also highlights that, due to the ability to teleport in VR, VR sound design may also differ not only from real-life sounds but also from sounds for other related technologies such as AR (augmented reality) and MR (mixed reality).

Sound Design Guidelines: When asked about any standard sound design guidelines, designers mentioned that some standards exist for technical aspects of the audio, such as file-formats (*e.g.,* raw or compressed), audio profiles (*e.g.,* HRTF [53]), channels (stereo, mono, surround, binaural, and 3D), broadcast standards [11] (*e.g.,* LKFS, LUFS), volume standards, and frequency standards. However, when asked about any creative design standards for VR audio (*e.g.,* how to design for multiple sounds, how best to position audio in 3D space), designers could not recall any. Since VR is a new technology, each designer has developed their own sound design practices:

> "I am not aware of any creative design standards—it's just from my own experience working with audio. [. . .] It feels like every product I am trying to come up with my own standard and my own guidelines and so there has never been any repeats because the VR platform has been developing." (D2)

The varied sound design practices imply that designing scalable visual/haptic feedback for sounds in VR may be difficult unless developer-specific customization options are incorporated.

Accessibility of VR Apps: Finally, we asked if designers had incorporated or thought about accessibility for their VR apps. Three designers had worked on specialized VR apps for people with visual impairment, two mentioned working in a team that designed a modified controller for people with limited mobility, and D6 had added a feature to customize the number of overlapping sounds in a mainstream app for users with cognitive disabilities. However, no participants have worked on accessibility features for DHH users. When asked for reasoning, D4 mentioned that as a sound designer, she tends to focus on accessibility for visually impaired people, and that accessibility for DHH users is *"maybe the job for visual artists who are visually oriented people."*

However, since deafness occurs on a spectrum [6], sound designers can indeed offer some customization features for people who have partial hearing such as frequency modulation or volume attenuation. When this was mentioned to the designers, they responded positively and asked for more guidelines on adding accessibility features for DHH people. For example:

> "You bring up a good point. [...] I wish there were more guidelines for sound designers to understand what things we should be careful of. For example, as you brought up, I would love to limit my frequency range or offer volume customization options, so it is more accessible. I wish somebody had told me this before. . .." (D4)

Discussion: Our findings from interviews with 10 VR sound designers reinforce the need to support sound accessibility in VR for DHH people. To enable the design of future accessible solutions, we characterized how sounds are used in VR through a novel sound taxonomy. In contrast to past work on movie and game sounds [14, 15], our taxonomy articulates both the source of the sound and the intent of the sound, which allows capturing of nuanced sound meaning (*e.g.,* a character's speech (a sound source) could convey critical information or could be a street noise for increasing realism). We also presented the difference between VR and real-life sound design to assess the extent to which real-life sound feedback designs (*e.g.,* [21, 30]) might be leveraged for VR. For example, if a VR experience involves teleportation, the soundscape will not be continuous, and the time-series visualizations (*e.g.,* waveforms) may not work. Similarly, because VR involves aesthetic sounds (*e.g.,* Foley sounds [52]), using a real-life sound classifier (*e.g.,* [24, 25]) in VR may be infeasible as the classifier may not accurately identify all sounds. Moreover, we found that designers have varied approaches to VR sound design. Thus, designing scalable prototypes for sound accessibility may be difficult, and these prototypes may require developer-specific customization. On the other hand, some technical aspects of VR sound design (*e.g.,* file formats) have standardized representations, which will help with prototype scalability (*e.g.,* a visual waveform generator need only work with standard file formats). Next, we present our evaluation of the taxonomy.

4 STUDY 2: EVALUATION OF THE TAXONOMY

To evaluate our taxonomy of *sound source* and *sound intent* and to assess how sounds are used across different VR apps, we performed an analysis of sounds in 33 VR apps.

4.1 Method

Procedure: To incorporate a diversity of apps, we selected 10 distinct VR app categories from the Oculus store [62] (Table 2). For each category, we selected five apps (total 50). We then recruited eight hearing HCI researchers through social media. These researchers pre-owned a VR headset: a recruitment criterion set by our team since COVID-19 social distancing requirements prevented people from using shared equipment or visiting our lab. Each researcher selected a non-overlapping set of 3-5 apps of their choice from the list of 50 Oculus store apps and used each app for about 30 minutes. A spreadsheet template was used to document any sounds encountered while using the app by pausing the app after every five minutes or after the end of an event (*e.g.,* at the end of a shooting round). The spreadsheet also enabled participants to categorize each sound within our sound taxonomy as well as report on whether a

Table 2: The 33 VR apps in 10 categories analyzed for sounds in Study 2.

Social	Sports & Fitness	Music & Rhythm	Shooting Games	Racing Games	Puzzle Games	Relax & Meditate	Travel & Discovery	Art & Creativity	Movies & Media
AltSpace	Creed	Dance Central	SuperHot	Rush	Minecraft	Tripp	Star Chart	Tilt Brush	FireFox VR
Puppet Fever	Premium Bowling	Beat Saber	Robo Recall	Project Cars	Vacation Simulator	Guided Mediation	Ocean Rift	Kingspray Graffiti	The Under Presents
SculptrVR	Eleven: Table Tennis VR	Thumper	Arizona Sunshine	Ultrawings	Keep Talking & Nobody Explodes	Where Thoughts Go	Apollo 11	Pottery VR	Spice & Wolf VR
			Dead & Buried				Google Earth		
			Half life						

sound was accompanied by a visual and/or haptic feedback, whether it was spatialized, and any open-ended comments on noteworthy characteristics of a sound (*e.g.,* persistent *vs.* a discrete sound). In total, participants analyzed 33 VR apps (Table 2).

Data analysis: We performed thematic analysis on the open-ended comments and summarized other responses using descriptive statistics. For thematic analysis, we followed a similar procedure as described earlier for Study 1. One team member read the spreadsheet and created 13 open codes to summarize the data. The team member then reorganized the open codes to generate seven axial codes, which were further combined into three overarching themes. The themes and codes were verified with other team members and sound examples were collected for each code. For other responses, we used automatic spreadsheet functions to calculate the number of sounds of each category from our taxonomy, percentage of sounds in each category that were accompanied by visual/haptic feedback, and percentage of spatialized sounds within each category.

4.2 Findings

We summarize the taxonomizing process and the specific results based on sound source and intent.

Taxonomizing process: Participants reported a total of 267 sounds in 33 VR apps. When fitting these sounds to our taxonomy, nearly all sounds (265/267) were categorized using the "sound source" dimension; the two uncategorized sounds were: *"an ambient sound during the auction process"* (P1, *Pottery VR*) and *"a whoosh sound when the app starts"* (P8, *SculptrVR*). On following up, P1 reported that that they were unclear on whether the auction sound belonged to "surrounding ambience" (diegetic) or "music" (non-diegetic), and P8 was confused between "notification sounds" and "surrounding ambience" for the whoosh sound; this reveals a potential uncertainty in our taxonomy.

For the "sound intent" dimension, 41/267 sounds (15.4%) were not categorized (*i.e.,* marked as "other"). We performed a post-hoc investigation of the uncategorized sounds, finding that these sounds reflected another possible category in VR—that is, interactions that are not critical to game progression such as "picking up a decorative object." Thus, we revised the taxonomy by adding another category to this dimension—that is, "sounds for non-critical interaction" (Table 3), and asked participants to reanalyze the VR apps. Subsequently, all sounds were successfully categorized.

Results for "sound source" dimension: For each sound category, we report on the total number of sounds, the type of apps the sounds were common in, and whether the sounds were accompanied by visual and/or haptic feedback. Overall, interaction sounds (*e.g.,* picking up an object, clicking a menu) (27.7%) were the most common followed by inanimate objects (*e.g.,* weapons, sports equipment) (18.7%), notification sounds (*e.g.,* low on ammunition) (13.1%), and surrounding ambience (*e.g.,* forest sounds, fan in a room) (13.1%) (Table 3).

Surprisingly, music (9.0%) was less reported overall, despite being present in nearly all apps (29/33). On investigating this, we realized that while our analysis reported on the number of different sounds of each category, it failed to account for temporal spans of these sounds. Specifically, the counts of music sounds were low because they were of lesser variety, despite a greater temporal span. This is also corroborated by P4: *"music was very central to this app to simulate [a] meditation experience. But for the most part I think there was only [one] background music clip."*

To assess if the sound information could be accessed through other ways for DHH users, we also asked about whether a sound was accompanied by a complementary visual or haptic feedback. For visual cues, interaction sounds, point ambience, and object-based sounds were—by definition—always visible by their location in the space (89.8%) unless they were outside the field-of-view (10.2%). Critical notification sounds were also commonly accompanied by a visual cue (71.4%). For example, *"when you knock out your opponent, there is a high-pitched ambient sound and red streaks. . ."* (P5, *Creed*).

In contrast, surrounding ambience (17.1%) and music sounds (13.3%) were rarely visible except when a subtle visual indication was shown such as *"leaves rustling"* (P7) or *"crickets chirping"* (P4), or in music-based instruction apps, such as *Dance Central*, where an instructor *"dancing to the rhythm of the music"* (D1) is shown.

Moreover, while speech sounds were rarely transcribed, the amount of complementary visual cues was higher than expected (localized speech: 58.8%, non-localized speech: 50.0%) because, for this category, participants reported on any cues that gave an indication that a character was speaking, such as *"face movement"*

Table 3: Results of Study 2 for our sound taxonomy, detailing the percentages of each sound type in the 33 VR apps (%Count), and the percentage of sounds within each type that are accompanied by a visual feedback (%Visual), haptic feedback (%Haptic), both (%Both), or any (%Any).

Sound Source	%Count	%Visual	%Haptic	%Both	%Any	Sound Intent	%Count	%Visual	%Haptic	%Both	%Any
Localized speech	6.4	58.8	0.0	0.0	58.8	For critical information	37.8	63.4	2.0	21.8	87.1
Non-localized speech	2.2	50.0	0.0	0.0	50.0	Increasing realism	26.2	44.3	2.9	12.9	60.0
Animate objects	3.4	66.7	11.1	0.0	77.8	Rhythm or movement	2.2	100.0	0.0	0.0	100.0
Inanimate objects	18.7	60.0	4.0	22.0	86.0	Affective state	4.9	15.4	0.0	0.0	15.4
Notification sounds	13.1	71.4	2.9	22.9	86.0	Aesthetics or decoration	13.5	30.6	0.0	2.8	33.3
Interaction sounds	27.7	59.4	1.4	24.6	97.1	Non-critical interaction	15.4	53.7	4.9	17.1	75.6
Surrounding ambience	13.1	17.1	2.9	5.7	25.7						
Point ambience	6.4	79.4	0.0	5.9	85.3						
Music	9.0	13.3	0.0	0.0	13.3						

(P3) or *"gestures"* (P1). But only one of the apps (*Guided Meditation*) displayed transcription for some speech: *"the person was asking me which meditation to try and the same instructions were also [shown as text] on the top of the screen." (P4).*

In terms of haptics, sounds were rarely augmented by a high-attentional haptic feedback unless they were important, such as critical notification sounds (*e.g.,* low on ammunition, crossing a finish line), weapon sounds in shooting games, or critical interaction sounds. For example, *"on bomb blast, my controllers vibrate for a couple seconds."* (P1, *Keep Talking and Nobody Explodes*).

Results for "sound intent" dimension: As expected, sounds for conveying critical information (*e.g.,* enemy footsteps, low ammunition) (37.8%) and increasing realism (*e.g.,* river flowing, bird chirping) (26.2%) were the most commonly present (Table 3). Because of their importance, critical information sounds were also mostly accompanied by a redundant visual or haptic modality (87.1%) (*e.g.,* "birds can be seen" – P5, "smoke from tires screeching" – P8). For sounds that increase realism, two categories emerged: spatialized point ambience, which were mostly shown visually, and non-spatialized surrounding ambience, which were rarely shown.

In closing, while useful in some cases, we emphasize that presence of a complementary visual or haptic modality does not necessarily make an app fully accessible—as is corroborated by P1 and P8:

> "Ball is [visually] seen to be bouncing on the surface or hitting on a player's bat, but audio helps a lot when playing fast. It helps quickly recognize when the opponent has hit the ball, and the intensity of the [opponent's] shot. Then, I can plan my next actions beforehand. This is critical for fast playing games." (P1, *Eleven Table Tennis VR*).
>
> "I can see smoke from tire screeching [a visual feedback] but the loud uncomfortable screeching sound itself is the main thing that tells me, you know, that I need to steer my car right, otherwise it may break. . ." (P8, *Project Cars*).

Discussion: Our revised taxonomy was able to classify all 267 sounds reported in the 33 VR apps within the sound intent dimension and nearly all sounds (265/267) within the sound source dimension. We also found that interaction sounds, inanimate objects, and notification sounds were the most commonly present for the sound source dimension, and that critical information sounds and realism sounds were common for the sound intent dimension. As DHH individuals have varying hearing levels [6, 27], all sound types should be considered for designing accessible visual/haptic augmentations. Nevertheless, our findings show that some sounds may be more vital to the VR experience than others and should be prioritized. For example, critical information sounds—such as for notification purposes—are vital to progression in a VR app and need to have an accessible representation. Similarly, surrounding ambience sounds are rarely augmented by an alternative feedback modality and could be important to represent accessibly for individuals with partial hearing.

5 DISCUSSION

Too often, accessibility is only considered as an afterthought, resulting in inaccessible or sub-par user experiences [51]. As Mott *et al.* point out in their recent position paper [34], VR technologies are at a crossroads in time where there is still an opportunity to codify accessibility best-practices for this emerging medium. While researchers have begun to consider making VR accessible to those with diverse visual [46, 49, 55] and motor [38] abilities, the needs of DHH users in VR are as-yet-unexplored. In this work, we have presented the first comprehensive look at sound in VR with an eye toward supporting sound accessibility for DHH end-users. Our contributions include (1) a novel taxonomy that articulates both the source and intent of sounds in VR (and can be useful for DHH accessibility), and (2) empirical insights on sound design and sound accessibility in VR. Here, we discuss limitations and further implications of our work.

5.1 Towards VR Sound Accessibility

The two dimensions of our taxonomy (sound source and intent, Table 3) as well as the VR sound characteristics uncovered from Study 1 (identity, spatial location, loudness, pitch, persistence, priority, and environmental effects) can be leveraged to develop visual and haptic sound feedback for DHH users. Future work can use our findings to scaffold investigation of interfaces that map different sound categories to different forms of feedback (*e.g.,* ambient sounds can be conveyed using subtle haptic sensations on the body, object sounds can be notified using visual animations). A key design consideration will be avoiding cognitive overload, particularly when visual and haptic feedback are jointly delivered (*e.g.,* in a busy VR scene). One possibility is to use high-attentional haptic feedback to deliver low-bandwidth sound alerts (*e.g.,* loudness) and high-bandwidth visual feedback to convey additional sound information (*e.g.,* identity).

Besides designing visual and haptic substitutions, sound itself can be attenuated for accessibility. Deafness occurs on a spectrum [6] and many DHH people would benefit from features such as customizing the frequencies of VR sounds (to accommodate varying frequency hearing loss), independently customizing the volume of foreground and background sounds, and background noise cancellation.

5.2 Applications for Other Domains

While we targeted accessibility for DHH users, our work also has applications for other domains. For example, our taxonomy can help sound designers organize sounds during production, invent new sound effects, and design new sonic interactions for less explored sounds (*e.g.,* rhythm and movement, point ambience; see Table 3). Our findings on accessible VR representations can also benefit users who are hearing or have other disabilities, particularly in cases of situational impairments [44] and cognitive overload (*e.g.,* when using a noisy VR app for extended periods).

5.3 Further Evaluations

Our findings on sound accessibility in VR are based on qualitative studies with sound designers and HCI researchers. While valuable, future work should extend these findings by conducting scenario-based controlled studies with both hearing and DHH users (*e.g.,* [56]) to determine how the original VR experience (*e.g.,* immersion, game challenge) changes for DHH users. We could not conduct such a study because the prevalence of VR headset ownership is low (and even lower among the DHH population due to inaccessibility) and COVID-19 social distancing prevented us from conducting in-person evaluations. Still, our work was informed by the VR experiences of our two DHH paper authors (including the lead researcher), and future work should continue to involve the DHH population in different stages of research.

5.4 Limitations

We acknowledge our work may not be desired by all DHH people, since not all DHH people want sound feedback. At the same time, we argue that the DHH community is broad [6] and past large-scale surveys with DHH people [2, 13] as well as the experiences of our DHH authors suggest that many DHH people do appreciate sound information. Nevertheless, future work should also investigate non-sound related accessibility features for DHH users such as background blurring to focus visual attention [63].

Furthermore, while our taxonomy was able to conceptualize VR sounds accurately and semantically across a variety of app categories, we do not claim that it is exhaustive or the only way of categorizing sounds in VR. Indeed, as our findings show, VR sound design technology is still in its infancy, and as technologies evolve, so should related ontologies. We welcome future work that refines, extends, or reimagines our taxonomy. One potential improvement area is to investigate a better separation of speech *vs.* music sounds (e.g., a character singing could arguably be classified as either, although our Study 2 participants categorized all singing sounds as speech).

6 CONCLUSION

Ensuring that mainstream VR applications are accessible to people with a spectrum of hearing capabilities is an important and largely unexplored research challenge [34]. In this paper, we used interviews with sound designers and analysis of *status quo* VR apps to formally characterize sounds and sound accessibility in VR as well as lay a groundwork for progress toward approaches for accessible sound representations in this emerging medium. Our work advances sound accessibility in VR by articulating a novel taxonomy of VR sounds and providing empirical insights into sound design in VR.

ACKNOWLEDGMENTS

We thank Ed Cutrell, Nikunj Raghuvanshi, Travis Fodor, and Ivan Tashev for their valuable feedback on our work.

REFERENCES

[1] Chris Bevan, David Philip Green, Harry Farmer, Mandy Rose, Kirsten Cater, Danaë Stanton Fraser, and Helen Brown. 2019. Behind the curtain of the" ultimate empathy machine" on the composition of virtual reality nonfiction experiences. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–12.

[2] Danielle Bragg, Nicholas Huynh, and Richard E. Ladner. 2016. A Personalizable Mobile Sound Detector App Design for Deaf and Hard-of-Hearing Users. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, 3–13.

[3] Alexander Brandon. 2004. *Audio for Games: Planning, Process, and Production (New Riders Games)*. New Riders Games.

[4] Felix Burkhardt, Astrid Paeschke, Miriam Rolfes, Walter F Sendlmeier, and Benjamin Weiss. 2005. A database of German emotional speech. In *Ninth European Conference on Speech Communication and Technology*.

[5] Stuart K Card, Jock D Mackinlay, and George G Robertson. 1990. The design space of input devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 117–124.

[6] Anna Cavender and Richard E Ladner. 2008. Hearing impairments. In *Web accessibility*. Springer, 25–35.

[7] David J Chalmers. 2017. The virtual and the real. *Disputatio* 9, 46: 309–352.

[8] Michael Chion. 1994. Audio-vision: Sound on Screen Columbia University Press. *New York*.

[9] Les Christidis and Walter Boles. 2008. *Systematics and taxonomy of Australian birds*. Csiro Publishing.

[10] Juliet M Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology* 13, 1: 3–21.

[11] R EBU–Recommendation. 2011. Loudness normalisation and permitted maximum level of audio signals.

[12] Sigal Eden and Sara Ingber. 2014. Enhancing storytelling ability with virtual environment among deaf and hard-of-hearing children. In *International Conference on Computers for Handicapped Persons*, 386–392.

[13] Leah Findlater, Bonnie Chinh, Dhruv Jain, Jon Froehlich, Raja Kushalnagar, and Angela Carey Lin. 2019. Deaf and Hard-of-hearing Individuals' Preferences for Wearable and Mobile Sound Awareness Technologies. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)*.

[14] Troels Folmann. 2004. Dimensions of game audio. Unpublished. Available at http://www. itu. dk/people/folmann/2004/11/dimensions-of-game-audio. html [Accessed October 28, 2005.].

[15] Johnny Friberg and Dan Gärdenfors. 2004. Audio games: new perspectives on game audio. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, 148–154.

[16] Jort F Gemmeke, Daniel P W Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio set: An ontology and human-labeled dataset for audio events. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 776–780.

[17] Kathrin Gerling and Katta Spiel. 2021. A Critical Examination of Virtual Reality Technology in the Context of the Minority Body. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York*, NY, USA.

[18] Wendell Hanna. 2007. The new Bloom's taxonomy: Implications for music education. *Arts Education Policy Review* 108, 4: 7–16.

[19] Sander Huiberts. 2010. Captivating sound the role of audio for immersion in computer games. University of Portsmouth.

[20] Dhruv Jain, Bonnie Chinh, Leah Findlater, Raja Kushalnagar, and Jon Froehlich. 2018. Exploring Augmented Reality Approaches to Real-Time Captioning: A Preliminary Autoethnographic Study. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*, 7–11.

[21] Dhruv Jain, Leah Findlater, Christian Volger, Dmitry Zotkin, Ramani Duraiswami, and Jon Froehlich. 2015. Head-Mounted Display Visualizations to Support Sound Awareness for the Deaf and Hard of Hearing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 241–250.

[22] Dhruv Jain, Rachel Franz, Leah Findlater, Jackson Cannon, Raja Kushalnagar, and Jon Froehlich. 2018. Towards Accessible Conversations in a Mobile Context for People who are Deaf and Hard of Hearing. In *Proceedings of the International Conference on Computers and Accessibility (ASSETS)*.

[23] Dhruv Jain, Angela Carey Lin, Marcus Amalachandran, Aileen Zeng, Rose Guttman, Leah Findlater, and Jon Froehlich. 2019. Exploring Sound Awareness in the Home for People who are Deaf or Hard of Hearing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 94:1-94:13.

[24] Dhruv Jain, Kelly Mack, Akli Amrous, Matt Wright, Steven Goodman, Leah Findlater, and Jon E Froehlich. 2020. HomeSound: An Iterative Field Deployment of an In-Home Sound Awareness System for Deaf or Hard of Hearing Users. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), 1–12.

[25] Dhruv Jain, Hung Ngo, Pratyush Patel, Steven Goodman, Leah Findlater, and Jon Froehlich. 2020. SoundWatch: Exploring Smartwatch-based Deep Learning Approaches to Support Sound Awareness for Deaf and Hard of Hearing Users. In *ACM SIGACCESS conference on Computers and accessibility*, 1–13.

[26] Kristine Jørgensen. 2011. Time for new terminology?: Diegetic and non-diegetic sounds in computer games revisited. In *Game sound technology and player interaction: Concepts and developments*. IGI Global, 78–97.

[27] Ines Kožuh, Manfred Hintermair, and Matjaž Debevc. 2016. Community building among deaf and hard of hearing people by using written language on social networking sites. *Computers in Human Behavior* 65: 295–307.

[28] Paddy Ladd and Harlan Lane. 2013. Deaf ethnicity, deafhood, and their relationship. *Sign Language Studies* 13, 4: 565–579.

[29] PerMagnus Lindborg. 2016. A taxonomy of sound sources in restaurants. *Applied Acoustics* 110: 297–310.

[30] Tara Matthews, Janette Fong, F. Wai-Ling Ho-Ching, and Jennifer Mankoff. 2006. Evaluating non-speech sound visualizations for the deaf. *Behaviour & Information Technology* 25, 4: 333–351.

[31] Matthew S Moore. 1992. For Hearing people only: Answers to some of the most commonly asked questions about the Deaf community, its culture, and the" Deaf Reality". Deaf Life Press.

[32] Tsuyoshi Morimoto, Noriyoshi Uratani, Toshiyuki Takezawa, Osamu Furuse, Yasuhiro Sobashima, Hitoshi Iida, Atsushi Nakamura, Yoshinori Sagisaka, Norio Higuchi, and Yasuhiro Yamazaki. 1994. A speech and language database for speech translation research. In *Third International Conference on Spoken Language Processing*.

[33] Meredith Ringel Morris, Jazette Johnson, Cynthia L Bennett, and Edward Cutrell. 2018. Rich representations of visual content for screen reader users. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–11.

[34] Martez Mott, Edward Cutrell, Mar Gonzalez Franco, Christian Holz, Eyal Ofek, Richard Stoakley, and Meredith Ringel Morris. 2019. Accessible by design: An opportunity for virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 451–454.

[35] Douglas A Nelson, Sandra L L Gaunt, C L Bronson, and T J Kloth. 2001. Database design for an archive of animal sounds. *IEEE Engineering in Medicine and Biology Magazine* 20, 3: 76–80.

[36] S C W Ong and S Ranganath. 2005. Automatic sign language analysis: a survey and the future beyond lexical meaning. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27, 6: 873–891.

[37] François Pachet, Daniel Cazaly, and others. 2000. A taxonomy of musical genres. In *RIAO*, 1238–1245.

[38] Shanmugam Muruga Palaniappan, Ting Zhang, and Bradley S Duerstock. 2019. Identifying Comfort Areas in 3D Space for Persons with Upper Extremity Mobility Impairments Using Virtual Reality. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 495–499.

[39] Phil Parette and Marcia Scherer. 2004. Assistive Technology Use and Stigma. *Education and Training in Developmental Disabilities-September 2004*: 217–226.

[40] David Passig and Sigal Eden. 2001. Virtual reality as a tool for improving spatial rotation among deaf and hard-of-hearing children. *CyberPsychology & Behavior* 4, 6: 681–686.

[41] Yi-Hao Peng, Ming-Wei Hsu, Paul Taele, Ting-Yu Lin, Po-En Lai, Leon Hsu, Tzu-chuan Chen, Te-Yen Wu, Yu-An Chen, Hsien-Hui Tang, and Mike Y. Chen. 2018. SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)*, Paper No. 293.

[42] Marti L Riemer-Reiss and Robbyn R Wacker. 2000. Factors associated with assistive technology discontinuance among individuals with disabilities. *Journal of Rehabilitation* 66, 3.

[43] J Salamon, C Jacoby, and J P Bello. 2014. A Dataset and Taxonomy for Urban Sound Research. In *22nd {ACM} International Conference on Multimedia (ACM-MM'14)*, 1041–1044.

[44] Andrew Sears, Min Lin, Julie Jacko, and Yan Xiao. 2003. When computers fade: Pervasive computing and situationally-induced impairments and disabilities. In *HCI international*, 1298–1302.

[45] Kristen Shinohara and JO Wobbrock. 2011. In the shadow of misperception: assistive technology use and social interactions. *SIGCHI Conference on Human Factors in HCI*: 705–714.

[46] Alexa F Siu, Mike Sinclair, Robert Kovacs, Eyal Ofek, Christian Holz, and Edward Cutrell. 2020. Virtual Reality Without Vision: A Haptic and Auditory White Cane to Navigate Complex Virtual Worlds. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.

[47] Axel Stockburger. 2003. The game environment from an auditive perspective. *Level Up*: 4–6.

[48] Mauro Teófilo, Alvaro Lourenço, Juliana Postal, and Vicente F Lucena. 2018. Exploring virtual reality to enable deaf or hard of hearing accessibility in live theaters: A case study. In *International Conference on Universal Access in Human-Computer Interaction*, 132–148.

[49] Ryan Wedoff, Lindsay Ball, Amelia Wang, Yi Xuan Khoo, Lauren Lieberman, and Kyle Rector. 2019. Virtual showdown: An accessible virtual reality game with scaffolds for youth with visual impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15.

[50] Elisabeth Weis and John Belton. 1985. *Film sound: Theory and practice*. McGill-Queen's Press-MQUP.

[51] Jacob O Wobbrock, Shaun K Kane, Krzysztof Z Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-Based Design: Concept, Principles and Examples. *ACM Transactions on Accessible Computing* 3, 3: 9:1–9:27.

[52] Benjamin Wright. 2014. Footsteps with character: the art and craft of Foley. *Screen* 55, 2: 204–220.

[53] Bosun Xie. 2013. *Head-related transfer function and virtual auditory display*. J. Ross Publishing.

[54] Alina Zajadacz. 2015. Evolution of models of disability as a basis for further policy changes in accessible tourism. *Journal of Tourism Futures* 1, 3: 189–202.

[55] Yuhang Zhao, Cynthia L Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. 2018. Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–14.

[56] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D Wilson. 2019. SeeingVR: A set of tools to make virtual reality more accessible to people with low vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14.

[57] Franz Zotter and Matthias Frank. 2019. Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality. Springer Nature.

[58] Home | XR Access Initiative. Retrieved September 6, 2020 from https://xraccess.org/

[59] XR Accessibility User Requirements. Retrieved September 6, 2020 from https://www.w3.org/TR/xaur/

[60] Spice & Wolf VR Leaves You Wanting For More Of Lawrence And Holo's Sweet Interactions. Retrieved September 8, 2020 from https://www.siliconera.com/spice-wolf-vr-leaves-you-wanting-for-more-of-

lawrence-and-holos-sweet-interactions/
[61] The Persistence for PS VR - PlayStation.Blog. Retrieved September 8, 2020 from https://blog.playstation.com/2018/10/11/the-persistence-for-ps-vr-gets-huge-free-update-october-18/
[62] Rift Store: VR Games, Apps, & More | Oculus.
[63] How Microsoft is using empathy to lead innovation – Innovation Stories.