SoundVizVR: Sound Indicators for Accessible Sounds in Virtual Reality for Deaf or Hard-of-Hearing Users

Ziming Li
zl1398@rit.edu
School of Information,
Rochester Institute of Technology
Rochester, New York, USA

Shannon Connell
sdcnai@rit.edu
National Technical Institute for the Deaf,
Rochester Institute of Technology
Rochester, New York, USA

Wendy Dannels
w.dannels@rit.edu
Center on Culture and Language,
National Technical Institute for the Deaf,
Rochester Institute of Technology
Rochester, New York, USA

Roshan Peiris
roshan.peiris@rit.edu
School of Information,
Rochester Institute of Technology
Rochester, New York, USA

ABSTRACT

Sounds provide vital information such as spatial and interaction cues in virtual reality (VR) applications to convey more immersive experiences to VR users. However, it may be a challenge for deaf or hard-of-hearing (DHH) VR users to access the information given by sounds, which could limit their VR experience. To address this limitation, we present "SoundVizVR", which explores visualizing sound characteristics and sound types for several types of sounds in VR experience. SoundVizVR uses Sound-Characteristic Indicators to visualize loudness, duration, and location of sound sources in VR and Sound-Type Indicators to present more information about the type of the sound. First, we examined three types of Sound-Characteristic Indicators (On-Object Indicators, Full Mini-Maps and Partial Mini-Maps) and their combinations in a study with 11 DHH participants. We identified that the combination of Full Mini-Map technique and On-Object Indicator was the most preferred visualization and performed best at locating sound sources in VR. Next, we explored presenting more information about the sounds using text and icons as Sound-Type Indicators. A second study with 14 DHH participants found that all Sound-Type Indicator combinations were successful at locating sound sources.

CCS CONCEPTS

• **Human-centered computing → Accessibility technologies**; *User studies.*

KEYWORDS

virtual reality, audio visualization, deaf and hard-of-hearing, accessibility

ACM Reference Format:
Ziming Li, Shannon Connell, Wendy Dannels, and Roshan Peiris. 2022. SoundVizVR: Sound Indicators for Accessible Sounds in Virtual Reality for Deaf or Hard-of-Hearing Users. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22), October 23–26, 2022, Athens, Greece.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3517428.3544817

1 INTRODUCTION

Figure 1: SoundVizVR method that uses Mini-Maps and On-Object Indicators to present sound source characteristics and sound type information. The Mini-Map visualizes sounds in the VR environment while the On-Object Indicator displays the sound originating from an object ("a text that says 'Bark' is shown"). This example shows using icons in the Full Mini-Map and text on the object to describe the sound type.

Virtual reality (VR) technologies have the capability to deliver a completely immersive experience to the user. This experience allows the user to completely engulf or immerse their senses in the content and the interactions. Immersive VR experiences are typically dependent on the quality of the visual, sound, and interaction dimensions of the experience [35, 38]. As such, sounds in VR that enhance and complement the interactions are a crucial part of the "immersive" virtual reality experience. VR utilizes sounds in many forms, such as spatial audio, voice, interaction sounds,

---

ASSETS '22, October 23–26, 2022, Athens, Greece

ACM ISBN 978-1-4503-9258-7/22/10…$15.00
https://doi.org/10.1145/3517428.3544817

rhythmic interaction, background sounds/music, etc., to enhance the interaction experience [18].

However, although sounds in VR play an essential role in creating the immersive experience, sounds could also be a limiting factor for deaf or hard-of-hearing (DHH) individuals to experience a fully immersive experience in VR due to the limited sound accessibility [19, 26]. To address similar limitations in everyday situations (in the non-VR space), visualization-based [17, 20, 24] or haptics-based [15, 29, 30] methods have been proposed to make everyday sounds accessible for DHH persons. However, adapting the above practices to the VR space presents a set of unique challenges due to the fictional nature of the visual world and novel interaction possibilities VR presents [25]. With these challenges in mind, a few pioneering research has already begun exploring making sounds in virtual reality more accessible for DHH users. Here, Jain et al. defined a "Taxonomy of Sounds in Virtual Reality" to help future VR designers design accessible sounds in VR [18]. In addition, Jain et al. also presented an exploration of the design space for accessible VR sounds using visual and haptic-based method [19] and another pioneering work, EarVR [26], presented a haptic-based method for presenting spatial sounds for DHH VR users.

Inspired by the above approaches, we present SoundVizVR that explores visualizing VR sounds via sound indicators. Specifically, SoundVizVR uses Mini-Map Indicators and On-Object Indicator style interfaces as Sound-Characteristic Indicators and uses text and icons as Sound-Type Indicators to increase the ability to localize sound sources and visualize sound characteristics in VR (Figure 1). Our proposed sound visualization methods aim to enable users to visualize several types of sounds that originate from the sound sources within the VR environments [18], by identifying sound characteristics (e.g., loudness, duration) and sound types (e.g., footsteps, gunshots). Influenced by Stockburger [42], we use the term *diegetic sound* for the sound that comes from the object in the VR world (e.g., a phone ringing). For the current scope of this work, we focus only on several diegetic sounds in VR [18, 42]: localized speech, inanimate objects, animate objects, and point ambience, as they are critical for the experience in VR [18]. We present our prototype as a generalizable and customizable plugin[1] that can be integrated into any VR software developed in the Unity game engine [2]. Using this Unity plugin, we envision that VR designers and developers can promptly visualize sound source information to increase the sound accessibility for DHH VR users.

We conducted two user studies to explore the usability of firstly, the *Sound-Characteristic Indicator* visualization method and secondly, the *Sound-Type Indicator* visualization method. In the first study, we conducted a preliminary evaluation with 11 DHH participants to identify the best and the most preferred Sound-Characteristic Indicator visualization technique to *localize* sound sources from six design combinations of the *Mini-Map* based and the *On-Object Indicator* based sound visualization method. These methods assisted DHH VR users in locating the sound sources and visualizing other characteristics of the sounds, such as the loudness and duration. The best technique of the first study was selected for the second study based on the performance data and participants' feedback. In the second study, we integrated the chosen visualization technique into a VR game scene to explore the best and most preferred type of indicator for presenting the *sound types* (e.g., footsteps, gunshots, etc.). We implemented combinations of *icons* and *text* representation methods into the visualization technique to present the different sound types. We conducted a user experiment with 14 DHH participants to collect user performance data and user feedback to evaluate these combinations.

In summary, our research contributions are as follows:

- A VR sound visualization prototype software that can be used as a Unity platform plugin to improve sound accessibility for DHH users in VR projects
- A study with 11 DHH participants that identifies characteristics of the combinations of Mini-Map Indicators and On-Object Indicators to represent sound characteristics in VR, and
- A study with 14 DHH participants that discusses the preferences of presenting sound types through icons and text in the Full Mini-Map and on virtual objects.

2 RELATED WORK

2.1 Sound Visualization

Prior work has mainly explored visual and haptic methods, aiming to make sounds more accessible for DHH individuals. Here, we focus on prior work closely related to our work that applied visual-based sound accessibility methods.

Sound visualization techniques used in music can convey music information (e.g., pitch, tempo, etc.) in real-time with animated images [41]. Music visualization methods are widely supported by media player software, such as Windows Media Player, Foobar2000, and iTunes. They visualize music characteristics using spectrum-like or waveform-like 2D displays. There is also research on music visualization techniques conducted among DHH communities to evaluate DHH people's experience with these methods. For example, S. Nanayakkara et al. implemented a system combining a vibrated chair with a visual display to provide an enhanced musical experience to DHH people [29]. In addition, J. Mori and D. I. Fels conducted research with DHH people to investigate their emotional reaction to a song with different animated lyrics [28], which indicated that the animated text could provide entertainment value of the music without losing the readability of lyrics.

Similarly, sounds also have vital contributions to the experience designs in games. There are video games that look into using sound visualization methods to make their content more accessible to DHH people. For example, a sandbox video game called "Minecraft" [27] features "Subtitles" that use text labels with arrows to indicate the sound types and the directions of in-game sounds (e.g., rain falls, zombie groans, etc.) near the player avatar [47]. And a battle royale game called "Fortnite" [7] uses a radar-like interface to assist DHH players in accurately locating the vital sound effects in the game environment, like footsteps and gunfires. However, although with caption supported [23], the playing experience is still significantly reduced in many commercial video games once sounds are disabled. This limitation may be due to the lacking sound accessibility designs in their major game events (e.g., notifications of shootings from an enemy in a first-person shooter [FPS] game) [3, 48].

---

[1] https://github.com/ZimingLiii/SoundVizVR-Plugin

[2] https://unity.com/

Inspired by these prior works, we are exploring implementing sound visualization methods that can provide sound characteristic information. Our proposed methods adopt the interface components of Mini-Maps and environmental indicators, influenced by the accessibility designs of existing video games.

2.2 XR Sound Accessibility

Previous research primarily addressed accessibility for people with visual impairments [40, 46, 50] or mobility impairments [9, 10, 32] using XR technologies (including virtual reality, augmented reality, etc.). For DHH users, researchers mainly investigated captioning, sound awareness, description of sounds, and the ability to locate sounds in XR environments.

Prior work on sound accessibility in augmented reality (AR) explores speech captioning for talking people using wearable AR devices [14, 31, 34]. For example, D. Jain et al. [14] evaluated real-time captioning in an AR approach, which provided insights on the UI design of speech captioning systems in AR head-mounted displays (HMDs) and outlined the benefits of AR head-mounted display captioning. Another work from Y. Peng et al. [34] proposed an AR captioning system called "SpeechBubbles" to address problems in the group conversation scenario, such as the out-of-view captions and the speaker association. Also, prior research in AR accessibility addressed sound awareness [16] and sound detection and localization [11] which provide glanceable sound information of the environment to DHH people.

In terms of VR, prior work mainly explores providing accessible acoustic clues using subtitles [1, 13, 36, 44]. Several other research examines VR in the context of story telling [4] for DHH individuals and learning sign language for DHH children [33]. And studies like EarVR [26] also looked into conveying sound information in VR through haptics. Some VR games in the market adopt sound accessibility interfaces other than captions, for example, the enemy sound direction indicator in "The Persistence" [22].

Our work is influenced by prior work from D. Jain et al., which proposed a sound taxonomy in VR [18]. And also, our study is inspired by another research that presented an evaluation of their prototypes developed using visual-based and haptic-based representation methods for VR sounds [19]. However, to the best of our knowledge, limited research examines sound visualization methods in VR in detail. Therefore, it would be beneficial to conduct user experiments to help gain qualitative and quantitative insights, such as performance, usability, and user experience, of the sound visualization method designs for VR.

3 SOUNDVIZVR SOUND VISUALIZATION SYSTEM

The scope of this work mainly focuses on diegetic types of sounds, specifically localized speech, inanimate objects, animate objects, and point ambience [18]. To visualize these sounds in VR, we focused on two aspects of the sound indicators - 1) Sound-Characteristic Indicator and 2) Sound-Type Indicator. The Sound-Characteristic Indicator was implemented to present information such as the location, loudness, and the duration of sounds. The Sound-Type Indicator was used to convey more meaningful information about the sounds, such as a footstep sound from a person walking nearby.

3.1 Sound-Characteristic Indicator System Design

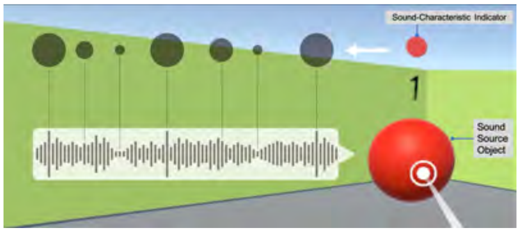

Figure 2: The design of the Sound-Characteristic Indicator. (Study 1 game scene is shown.) The smaller red object is the Sound-Characteristic Indicator which appears when the sound source object makes a sound and dynamically changes its size during the sound. The waveform in the speech bubble shows a sound being played, and the circles above indicate how the indicator changes the size based on the loudness characteristics of the sound wave.

The Sound-Characteristic Indicator was designed as a customizable 3D object (a Unity plugin). It is attached as a part of any sound source object in the VR design. The indicator appears in a Mini-Map and on the sounding object when the object starts making sounds and disappears at the end of the sound. As shown in Figure 2, it can visualize the current loudness of the sound by dynamically changing its size - when the sound gets louder, the indicator expands; when the sound is lower, the indicator shrinks. In this way, the dynamic changes in the object's size indicate the sound's loudness, and the appearance and disappearance of the object relate to the duration of the sound.

In addition, to indicate the location of the sounding objects and inform the loudness of the sounds, we designed two components for our proposed sound visualization system: Mini-Maps and On-Object Indicator.

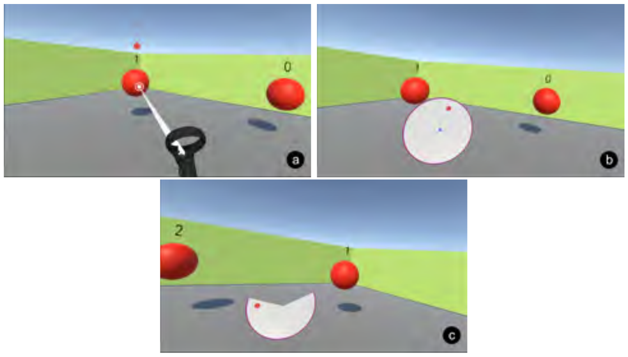

Figure 3: Three types of Sound-Characteristic Indicators: (a) On-Object Indicators; (b) Full Mini-Map Indicator; (c) Partial Mini-Map Indicator

3.1.1 On-Object Indicator. We wanted the user to be aware of the sound playing from a sounding object while looking at the object, which is an intuitive way to identify a sounding object. Hence, as seen in Figure 3a, we placed the Sound-Characteristic Indicator object hovering near the top of the object.

The appearance of the indicator is customizable. To make its design simple and visually apparent, we designed the appearance of the On-Object Indicator as a red-colored sphere for Study 1. However, designers may customize the indicator using any 3D objects or images based on their requirements.

3.1.2 Mini-Maps. The Mini-Map systems are commonly used in many video games [49]. The Mini-Map design can present players with information about their surroundings. So the player can keep track of the notable updates around them, especially those in their blind spots. Inspired by such map designs, we integrated a circular Mini-Map, which is called a "Full Mini-Map", into our proposed system to help users locate the sound source (Figure 3b). Through trial and error and feedback from pilot demos, we placed the Mini-Map on the screen's left side, similar to many video games. Furthermore, the same Sound-Characteristic Indicator described above is visualized to present sound source information on the Full Mini-Map.

In addition to the Full Mini-Map, we also present a Partial Mini-Map (Figure 3c). Here, the front sector of the Partial Mini-Map, which represents the user's current field of view, is hidden. We explored this design so that the user is only presented with sound information that is not in their field of view. In addition, this design aims to encourage users to pay more attention to the environment in front of them and observe potentially sounding objects instead of over-relying on the information shown on the Mini-Maps.

3.2 Sound-Type Indicator

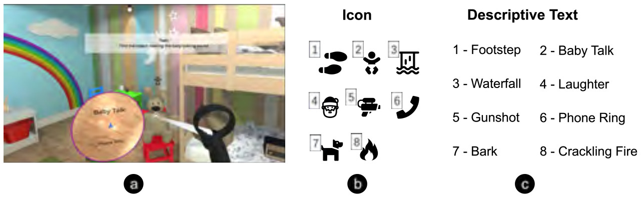

Figure 4: The design of Sound-Type Indicator. (a) The Sound-Type Indicator with text on the Full Mini-Map and an icon on the object. Here, in the player's view, the plushie rabbit makes a "talking sound", and its corresponding text and icon are shown. (b) Icons used for the presented sounds. (c) Text descriptions of the sounds.

Besides knowing the sound characteristics, it is essential to understand what types of sounds are presented in a scenario. Thus, to assist the DHH users in identifying the types of the sounds in the VR environment, we explored customizing the sound indicator in the On-Object Indicator and Mini-Maps with icons and text (Figure 4a). The size of the icon (Figure 4b) and text (Figure 4c) changes along with the loudness of the indicator's corresponding sound. In this way, a sound indicator can inform the sound characteristics to the user and further notify the sound type via its iconic or textual label at the same time.

3.3 System Implementation

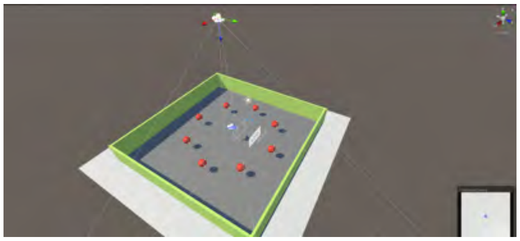

Figure 5: The implementation of a game scene developed using the Unity game engine.

We developed two task-based VR game scenes with our proposed system integrated using the Unity game engine version 2021.1.15f1. Both scenes can be deployed to Oculus's PC-powered VR headsets. In our case, an Oculus Rift S with dual VR controllers is used [6].

To implement the sound indicator, first, we need to obtain the sample data from the audio source of the current time frame. Then, we retrieve the average absolute value from the sample data and map it to the scale value of the sound indicator object of each frame. As a result, we can see the indicator's size changes according to the current loudness of the audio source (Figure 2) during the game engine's run-time.

In terms of the Mini-Maps' implementation, we use an additional virtual camera in the Unity game engine that is attached to the player at the ceiling level (Figure 5). This virtual camera always faces the ground. It ignores all other objects in the view except for the sound indicators. By rendering this virtual camera's view into the Mini-Map's texture, the Mini-Maps can reflect sound indicators around the player. After importing the plugin to the Unity editor, the designer could drag and drop the add-on prefab onto a sound-source object. Next, the designer would be able to customize the 3D objects, icons, and text based on the preferences or design requirements.

As mentioned above, the appearance of the sound indicator in the On-Object Indicator and Mini-Maps is customizable. Our prototype allows the designers to replace the default red sphere object with a preferred icon object or a descriptive text object for each sound source to further visualize the sound type information in a game scene.

4 STUDY 1: EVALUATING THE SOUND-CHARACTERISTIC INDICATOR

The main goal of the first study is to evaluate the performance and the user experience of our proposed Sound-Characteristic Indicator visualization methods on locating the sound source and visualizing sound characteristics like duration and loudness. For this purpose, we conducted a user experiment in which users were required to localize sound sources in VR as the main task to identify the best performing and the most preferred Sound-Characteristic Indicator method from the possible design combinations. Both studies presented in the article were approved by the institution's ethics review board.

4.1 Task Design

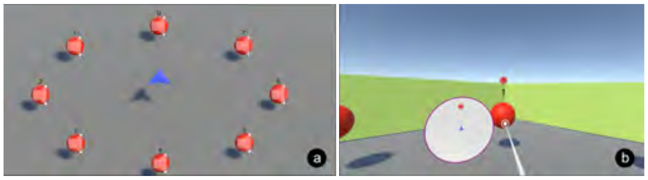

Figure 6: Study 1 task design - (a) VR environment of Study 1; (b) The participant's view of the scene while performing a task using the FM-OI visualization method.

Similar to the task design in EarVR [26], as the main task, the participant was required to locate the object (presented as spheres) from which the sound originated (Figure 6). In the experimental scene, we placed the player in the center of the stage with eight identical spheres evenly placed around the player in a circle. The spheres were labeled from 0 to 7 in counterclockwise order. The player could look around or lead to any direction but was not required to walk in the scene.

In the experiment, one of the eight spheres served as a sound source and started to play a random sound clip selected from an audio clip set. The audio clip set was formed based on four categories chosen from the taxonomy of sounds in VR proposed by Jain et al. that represented diegetic sounds [18]: Localized speech, Inanimate objects, Animate objects, and Point ambience. The selected audio clips were organized into two duration scopes: short duration sounds (a sound clip's length was shorter than 3 seconds) and long duration sounds (a sound clip's length was over 15 seconds). In total, eight sound clips were included in the set. The audio clip set used in the experiment is shown in Table 1. It was selected from the royalty-free sound clip website "SoundBible" [3] to represent a wide range of sounds from VR games. In addition, short sound clips were explored to identify the impact of the visualization techniques on locating quick sounds that may appear and disappear in the visualization.

Table 1: Study 1 sound selection

Category	Short Duration	Long Duration
Localize Speech	Old Man Laugh	Baby Talk
Inanimate Objects	Gun Shot	Phone Ringing
Animate Objects	Footstep	Barking Dog
Point Ambience	Fire Burning	Waterfall

In the task, the participant was asked to select the sounding sphere using a VR controller as a pointer. If the participant could not identify the sounding sphere after 7 seconds from the start of the sound, they were allowed to press a "skip" button on the VR controller to skip to the next task. As the localization accuracy rate, we calculated how many times a participant accurately selected a correct sounding sphere. The next random sphere started playing another random sound clip in 3 seconds after the participant chose a sphere or pressed the "skip" button. One repetition of a condition consisted of eight such tasks (eight different sounds from eight spheres) and each condition was repeated three times. The participant had the right to stop the experiment at any point of time.

Table 2: Study 1 conditions

Independent Variable		**On-Object Indicators**	
		Without	***With***
Mini-Map	***Without***	NON	OI
	Full	FM	FM-OI
	Partial	PM	PM-OI

4.2 Study Design

Study 1 used a within-subject evaluation design that consisted of two independent variables: On-Object Indicator (Without On-Object Indicator or With On-Object Indicator) and Mini-Map Type (Full Mini-Map, Partial Mini-Map, or Without Mini-Map). The dependent variables were sound source localization accuracy and the time of completion of each task. In addition, we recorded additional data, such as the head rotation angles. In total, there were six conditions (Table 2): NON (no sound visualization technique was used), OI (On-Object Indicators only-Fig. 3a), FM (Full Mini-Map technique only-Fig. 3b), FM-OI (a combination of Full Mini-Map technique and On-Object sound indicator-Fig. 4a), PM (Partial Mini-Map technique only-Fig. 3c), and PM-OI (a combination of Partial Mini-Map technique and On-Object sound indicator). The NON condition was included as the baseline condition in which no Sound-Characteristic Indicator was presented, similar to existing VR experiences. Each task was repeated 3 times. Each participant faced a total of 144 tasks during the experiment (8 tasks x 3 repetitions x 6 conditions).

4.3 Participants

We recruited 11 DHH participants for the study from the authors' institution (6 males, 2 females, 3 non-binary people; ages 18-45, Mean = 27.36, SD = 8.82). The participant group consisted of 4 deaf and 7 hard-of-hearing participants. Five of the participants had used VR devices before. The participants were recruited through flyers and word-of-mouth advertising in the institution. Each participant was paid $15 after completing the user experiment.

4.4 Procedure

After signing the informed consent form, the participant was given an introduction to the system and asked to fill out a demographic questionnaire. The information was provided to the participant through text and slides. However, a hard-of-hearing research team member used sign language to discuss additional details if and when necessary. Also, the participant was asked to take off the hearing aid, if there was one, to ensure the visualization was the focus in this controlled study. Next, the participant put on the VR headset and held a VR controller. Before a condition started, the participant was

---

[3] https://soundbible.com/

given sufficient time to get familiar with the VR device's control, the VR game scene, and the Sound-Characteristic Indicator interface based on the condition. As the condition started, the participant was required to perform the tasks with the corresponding combination of the Sound-Characteristic Indicator visualization technique of the condition. The order of the condition for each participant was assigned randomly. After completing a condition, the participant was asked to fill out a post-condition questionnaire. Then, the participant could take a 3-minute break if needed. After completing all six conditions, the participant was asked to fill out a post-test questionnaire. The experiment took approximately 75 minutes for each participant.

4.5 Results of Study 1

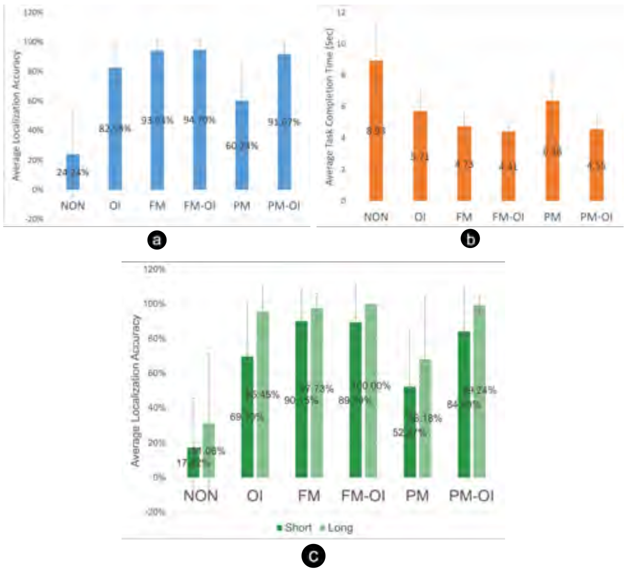

Figure 7: Results of Study 1 - (a) The average localization accuracy rate; (b) The average completion time (Sec) of each task; (c) The average localization accuracy rate on two types of sounds of Study 1. Error bars denote the standard deviation.

4.5.1 *Localization accuracy and completion time.*

Figure 7a shows the localization accuracy rate of Study 1. The results were further analyzed using repeated measures ANOVA with Greenhouse-Geisser correction. A significant main effect was found on Mini-Map Type ($F_{1.53,15.37}$ = 41.618, $p < 0.001$), and on On-Object Indicator ($F_{1,10}$ = 40.436, $p < 0.001$). Post-hoc comparison between each pairs of Mini-Map Type showed NON, FM: $p < 0.001$; NON, PM: $p < 0.001$; FM, PM: $p < 0.001$. Post-hoc comparison between each pairs of On-Object Indicator showed NON, OI: $p < 0.001$. Post-hoc comparison between each combination showed significant difference between each pairs, except FM, OI: $p = 0.540$; FM, FM-OI: $p = 1.000$; FM, PM-OI: $p = 1.000$; OI, FM-OI: 0.258; OI, PM-OI: 0.540; FM-OI, PM-OI: $p = 1.000$.

Figure 7b shows the task completion time of Study 1. The results were further analyzed using repeated measures ANOVA with Greenhouse-Geisser correction. A significant main effect was found on Mini-Map Type ($F_{1.591,15.912}$ = 27.453, $p < 0.001$), and on On-Object Indicator ($F_{1,10}$ = 31.983, $p < 0.001$). Post-hoc comparison between each pairs of Mini-Map Type showed NON, FM: $p < 0.001$; NON, PM: $p < 0.001$; FM, PM: $p = 0.029$. Post-hoc comparison between each pairs of On-Object Indicator showed NON, OI: $p < 0.001$. Post-hoc comparison between each combination showed significant difference between each pairs, except FM, OI: $p = 0.363$; FM, FM-OI: $p = 1.000$; FM, PM-OI: $p = 1.000$; PM, OI: $p = 0.843$; OI, FM-OI: $p = 0.085$; OI, PM-OI: $p = 0.146$; FM-OI, PM-OI: $p = 1.000$.

Figure 7c shows the localization accuracy rate on two types of sound duration of Study 1. The results were further analyzed using paired samples t-test. A significant main effect was found on all six conditions - NON: $p = 0.003$; OI: $p < 0.001$; FM: $p = 0.006$; FM-OI: $p = 0.003$; PM: $p = 0.006$; PM-OI: $p < 0.001$.

Table 3: The System Usability Scale Scores and Adjective Ratings of Study 1

Condition	SUS Score (SD)	Adjective Rating
NON	52.05 (21.12)	Poor
OI	71.14 (18.45)	Good
FM	84.55 (15.24)	Excellent
FM-OI	84.77 (16.03)	Excellent
PM	51.82 (23.69)	Poor
PM-OI	76.36 (24.91)	Good

4.5.2 *System usability and subjective mental workload.*

Table 3 shows the System Usability Scale (SUS) results of Study 1. The SUS adjective rating is obtained from the 7-point adjective scale proposed by Bangor et al. [2].

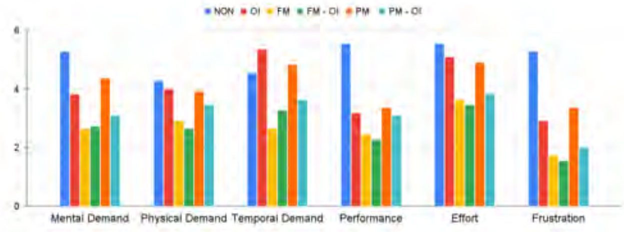

Figure 8: NASA TLX scores of Study 1. The lower ratings indicate lower task loads.

Figure 8 shows the subjective mental workload results collected from NASA Task Load Index (NASA TLX) questionnaires [12] across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration.

4.6 Discussion of Study 1

Overall, our results indicated that all the conditions with a Sound-Characteristic Indicator performed significantly better than the NON condition with FM-OI, FM, PM-OI achieving higher than 90% accuracy for sound source localization. The NON condition, which has no integrated sound visualization methods similar to the existing VR experiences, achieved the lowest localization accuracy (M:

24.24%, SD: 0.29) and the longest task completion time (M: 8.93s, SD: 2.21). During further analysis of the NON condition, we found that the correct localization was achieved by a few hard-of-hearing participants who, although, indicated that it was difficult for them to confidently locate the sound source without the assistance of visual cues. For example, one of the hard-of-hearing participants (F02) wrote in the feedback: *"I could not identify any of the sources of sound securely, I could maybe localize the sound to 3 spheres [three spheres that present in the view], but past that it's beyond me."* Similarly, the deaf participants reported that they could not either tell the location of sounds or the existence of a sound without any visualization. The NASA TLX ratings for the NON condition reported the highest ratings (higher ratings indicate more cognitive load requirements) for all categories except for the temporal demand category (Figure 8).

In terms of the FM condition, which introduced a Full Mini-Map Sound-Characteristic Indicator component compared to the NON condition, its localization accuracy reached 93.94% (SD: 0.09), and task completion reduced to 4.73s (SD: 0.89). It also outperformed the SUS score (with an "Excellent" adjective rating) and showed relatively low subjective mental workloads across all NASA TLX dimensions. It might indicate that the participants were able to identify the sound locations and perform tasks with the assistance of this method. The participants' feedback supported this observation. For example, F12 wrote: *"the Full Mini-Map helps me find the sound location more accurately."* Those who had a wide experience in playing games indicated that they might get used to the Full Mini-Map method quickly since the Full Mini-Map was commonly used in games [49]. For example, F03, who played games 4-6 times a week, wrote: *"I think that because I play video games with full circle Mini-Map on the screen, it is very easy for me to engage in this system."* However, some participants reported they had to rely on the map to perform tasks, even without paying attention to the environment. They also mentioned that, with the Full Mini-Map alone, they could fail to pinpoint the sound source from two close objects. For example, F02 added: *"I'm not entirely sure how accurate I was because I often had to choose between two of the spheres."*

The FM-OI condition, which was developed based on FM, combined a Full Mini-Map component with On-Object Sound-Characteristic Indicators that also supported visualizing the sound of an object in the environment. FM-OI had similar yet marginally better performance compared to FM on localization accuracy (M: 94.70%, SD: 0.07) and task completion time metrics (M: 4.41s, SD: 0.93), although not statistically significantly different. Similar to FM, FM-OI achieved an "Excellent" adjective rating in the SUS score and relatively low mental workloads across all six NASA TLX dimensions. The participants' comments reported that FM-OI enabled them to pay more attention to their surroundings with the assistance of the On-Object Indicators. *"Since the notification appears on both map and game, I was able to focus on the game and use the map as a support." (F01)* They reported that the Full Mini-Map component was useful to assist them in identifying the sound direction, and the On-Object Indicator component was helpful when identifying the specific sounding sphere. *"If there are two objects that are quite close to each other, knowing just the direction of the sound may not always be helpful. In such cases, having an environment sound indicator to differentiate between two objects in proximity would be an added advantage." (F14)*

In addition, we looked into OI to better understand the effect of the On-Object Indicator alone. Although there are no statistically significant differences when comparing its localization accuracy (M: 82.58%, SD: 0.15) and completion time (M: 5.71s, SD: 1.41) with FM and FM-OI, OI achieved significantly high subjective mental workloads in NASA TLX scores among all six dimensions compared to FM and FM-OI. The participant reported that they had to scan through all the objects to find the audio cues due to the lacking indicators of the sound direction. *"The problem with this system is although it does indicate the sound, it doesn't indicate the direction of the sound which could waste time in a high pressure gameplay." (F07)*

We looked into the results of PM and PM-OI to investigate if the Partial Mini-Map worked in reducing the information and enabled participants to focus on the environment as we intended. PM achieved the lowest localization accuracy (M: 60.23%, SD: 0.26) and longest task completion time (M: 6.38s, SD: 2.08) among the five proposed methods. In addition, the participants reported that it was difficult to identify the direction of the sounds with this method. *"If the sound was in front of me, I had no idea which of the 2 spheres it could be unless I looked to the side to see the dot on the map." (F02)* These kinds of feedback may back up PM's low SUS adjective rating ("Poor") and relatively high mental workload performance, especially in mental demand, temporal demand, effort, and frustration dimensions. In terms of the PM-OI condition, with an On-Object Sound Indicator component integrated, PM-OI achieved similar performance in localization accuracy (M: 91.67%, SD: 0.09) and task time completion (M: 4.55s, SD: 0.87) metrics compared to FM and FM-OI, which showed no statistically significant difference. The participants reported that the On-Object Sound Indicator enabled them to locate the sound source within multiple spheres on their front, which overcame the disadvantages of PM alone. *"The sound indicators help me find the sound location better when I can't find [it] with the Partial Mini-Map. If I [was] unable to locate the sounds with the front cut up, the sound indicators helped me find it faster and accurately." (F12)* Also, the participants reported that PM-OI could assist them in focusing more on the game itself. For example,

> "The dots over the top made things much easier, and having the radar showed what's around me made it clear when to turn. This is by far the easiest way to achieve the tasks. And yet the only time I missed the target. I learned there is no replacement, meaning that after one came and went, it would not be there again. That led me to reorient to the unselected options to get a 'jump' on them. Fun combination." (F08)

However, some participants indicated that the sound localization experience on PM-OI was ambiguous. As F07 said: *"I can't tell if I'm looking directly at the source of the noise in the Mini-Map with the notch and due to the nature of VR, I don't have the same field of vision as I would have in real life. So I have to put in a little extra effort to identify the sphere."* This may support PM-OI's higher subjective mental workload result of the Performance dimension in the NASA TLX ratings, especially when compared to FM and FM-OI.

When considering the duration of the clips used, sound localization in long-duration sounds had a significantly better localization accuracy than in short-duration sounds in all conditions. This result was expected here as we identified that if the participant was not in the view of a short duration sound, it had a higher chance of being missed in a majority of the conditions. Here, F01 indicated: *"For sounds that disappeared too quickly, it's difficult to notice if there were any sound at all. This is especially true for 'no map with indicator', 'Partial Mini-Map without indicator', and 'Partial Mini-Map with indicator'."* About the OI condition, F14 discussed: *"For this task, I had to scan through all the spheres until I found an audio indicator on top of one of the spheres. Sometimes, for short audio, I guess the audio indicator disappeared by the time I scanned through the spheres, which was not helpful in identifying the target... However, having an audio indicator was a lot better than having no visual cues at all!"* Also, it should be noted that we used different sounds for the different durations of the same type of sounds (we did not use a short and long version of the same clip) to represent a larger variety of sounds.

4.7 Summary of Study 1

Based on the analysis and observation of the results, we selected the FM-OI as the Sound-Characteristic Indicator visualization method for Study 2. We chose FM-OI over the six conditions primarily because it had the best performance data in both localization accuracy and task completion time. Similarly, the results from SUS, NASA TLX questionnaires, and participants' comments of Study 1 reported above supported our decision. Furthermore, in the post-test questionnaire, most participants (7/11) indicated that FM-OI was the sound visualization method they liked most among the five proposed methods, while preferences over the other methods were diverse (PM-OI: 2/11, OI: 1/11, FM: 1/11).

5 STUDY 2: EVALUATING THE SOUND-TYPE INDICATOR

Figure 9: VR environment of Study 2 - (a) An overview of the Study 2's scene. A participant stands at the location marked by the blue arrow; (b) The participant's view of the scene before a task. Before the task, the participant was shown the sound source objects with their labels displayed.

Based on the results from Study 1, we selected the FM-OI Sound-Characteristic Indicator visualization method for the following studies. As such, the main goal of Study 2 is to evaluate the performance and user experience of the different Sound-Type Indicator visualization methods (texts and icons).

5.1 Task Design

To present a more realistic VR scene for this experiment, we used the Kid's Room model package [5] and several additional object models [43, 45] from the Unity Asset Store to build our VR game scene for the study (Figure 9a). The participant was placed near center of the scene and the objects that serve as speakers were distributed in the scene. Similar to Study 1, we did not require the participants to move in the game scene. However, the participants could still rotate physically in place with minor lateral movements (rotate their bodies or lean towards a target) while wearing the VR headsets. Before a condition began, the participant was allowed to familiarize with the sound source objects that were indicated using a label showing the name (category) of the object (Figure 9b). In addition, we used the same set of sound clips from the previous study and used the same text label shown in Figure 4c. The used icons (Figure 4b) were selected from a free online icon database website called "Flaticon" [4]. To prevent any biases (due to preferences and/or color blindness), we used black and white icons.

Table 4: The task list of Study 2

No.	Task Instruction
1	Find the radio making the flowing water sound
2	Search for the item making a sound
3	Find and identify the ringing telephone
4	Find the object making the baby talking sound
5	Find and identify the sounding plush toy
6	Find the radio making the gunshot sound
7	Find the object making the campfire sound
8	Find the plushie making the Santa Claus laughter sound

Figure 10: Study 2 task design - (a) Task instruction panels; (b) The participant's view of the scene when performing a task. The TM-IO condition is shown.

As the main task, the participant was presented with a task from a list of task instructions (Figure 10a) as described in Table 4. It was presented in the main view for 7 seconds to ensure the participant had enough time to read it, and afterward, it was maintained in the upper region of the participant's VR view. Next, the VR scene played three different sounds from at most three sound source objects at the same time (Figure 10b). The participant was required to select the sound source object that played the sound specified in the instruction. We designed such a task to encourage the participant to focus more on the sound-type labels during the task. For example, during a given task, all three radios in the VR scene might make sounds, but if Instruction 1 was presented, the participant was required to identify the correct radio by looking for the "water flowing" icon or text label.

---

[4] https://www.flaticon.com/

The recording of task completion time began only at the point in time when the correct sound clip started to play from an object. As the localization accuracy rate, we calculated how many times a participant accurately selected a correct sounding object after the correct sound clip started to play.

Similar to Study 1, the participant was able to skip the task if they could not identify the correct task after 10 seconds by pressing the skip button. The participant also had the right to stop the experiment at any point of time. Eight such tasks (eight different sounds from eight objects) were included in each repetition of a condition. And each condition was repeated twice for a participant.

5.2 Study Design

The experiment of Study 2 used a within-subjects evaluation design. It consisted of two independent variables: Sound-Type Indicator on Objects (icon or text) and Sound-Type Indicator on the Full Mini-Map (icon or text). The dependent variables were sound source localization accuracy and the task completion time while searching for the correct sound source. Similar to Study 1, we recorded additional data such as the head rotation angles. In total, there were four conditions as shown in Table 5: IM-IO (Icon indicators on the Full Mini-Map + Icon indicators on Objects), IM-TO (Icon indicators on the Full Mini-Map + Text indicators on Objects), TM-IO (Text indicators on the Full Mini-Map + Icon indicators on Objects), and TM-TO (Text indicators on the Full Mini-Map + Text indicators on Objects). Each participant faced 64 tasks in total (8 tasks x 2 repetitions x 4 conditions).

Table 5: Study 2 conditions

Independent Variable		**Indicators on Objects**	
		Icon	***Text***
Indicators on Full Mini-Map	***Icon***	IM-IO	IM-TO
	Text	TM-IO	TM-TO

5.3 Participants

We recruited 14 DHH participants for the study from the authors' institution (8 males, 4 females, 2 non-binary people; ages 18-45, Mean = 26.21, SD = 8.31). The participant group consisted of 6 deaf and 8 hard-of-hearing participants. Six of the participants had used VR devices before. Ten of the participants had previously taken part in Study 1. However, it should be noted that the two research studies were conducted separately. The participants were recruited through advertising in the institution by flyers and word of mouth. Each participant was paid $25 after completing the user experiment.

5.4 Procedure

Similar to Study 1, after signing up for the informed consent form, the participant was given an introduction to the system and asked to fill out a demographic questionnaire. Next, after the participant put on the VR system, they were given sufficient time to familiarize themselves with the VR device's control and the VR game scene. In addition, the participants were allowed to familiarize themselves with the instructions and the name (category) labels and locations of the sound source objects. As the condition started, the participant was required to perform the tasks with the sound visualization technique of the condition. The order of the condition for each participant was assigned randomly. After completing a condition, the participant was asked to fill out a post-condition questionnaire. Then, the participant could take a 3-minute break if needed. After completing all four conditions, the participant was asked to fill out a post-test questionnaire. The experiment took approximately 45 minutes for each participant.

5.5 Results of Study 2

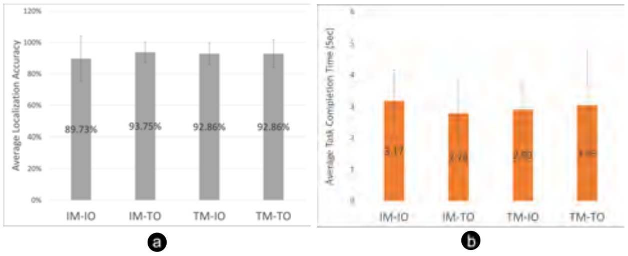

Figure 11: Results of Study 2 - (a) The average localization accuracy rate of each condition; (b) The average task completion time (Sec) of each condition. Error bars denote the standard deviation.

5.5.1 *Localization accuracy and task completion time.* Figure 11a shows the localization accuracy rate of Study 2. Repeated measures ANOVA revealed no significant difference on Indicators on Full Mini-Map ($F_{1,13}$ = 0.198, p = 0.664) and on Indicators on Objects ($F_{1,13}$ = 2.021, p = 0.179).

Figure 11b shows the task completion time of Study 2. Repeated measures ANOVA revealed no significant difference on Indicators on Full Mini-Map ($F_{1,13} = 6.596e - 4$, p = 0.980) and on Indicators on Objects ($F_{1,13}$ = 0.409, p = 0.553).

Table 6: The System Usability Scale Scores and Adjective Ratings of Study 2

	SUS Score (SD)	**Adjective Rating**
IM-IO	81.96 (14.01)	Excellent
IM-TO	73.39 (18.78)	Good
TM-IO	75.00 (15.81)	Good
TM-TO	76.25 (18.47)	Good

5.5.2 *System usability and subjective mental workload.* Table 6 shows the SUS results of Study 2. Figure 12 reveals the subjective mental workload results of Study 2 collected from NASA TLX questionnaires across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration.

5.6 Discussion of Study 2

5.6.1 *Sound Type Identification.* Overall, the quantitative results indicate that DHH participants can perform the tasks with the assistance of the four evaluated Sound-Type Indicator visualization methods and achieved high localization accuracy at around 90%.

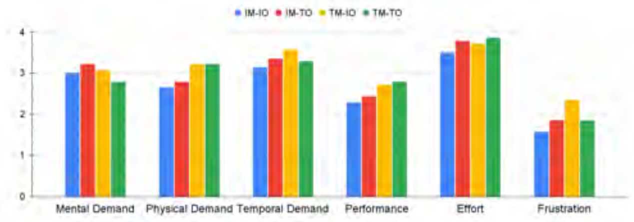

Figure 12: NASA TLX scores of Study 2. The lower ratings indicate lower task loads.

Based on the analysis of the results, there is no significant difference among all tested Sound-Type Indicator visualization methods on both localization accuracy and task completion time metrics of Study 2. It indicates that the overall performance of the four tested Sound-Type Indicator visualization methods are similar.

The icons used in Study 2 were selected based on what a designer would pick as icons for the user interface applying a nomic sound mapping method [8, 21]. Prior to the study, we did not present the icons used in the experiment and their descriptive text to the participants. It aimed to explore if the participants could identify the unfamiliar icons shown during the experiment. In the post-test questionnaire, we presented two five-point Likert scale questions to respectively investigate if participants could understand the meanings of icons and text shown in the indicator. The results show that the icons (M: 3.93, SD: 0.83) and text (M: 3.93, SD: 0.92) have similar ratings, which indicates the participants can understand the meanings of both the icons and text. The results of the system usability evaluation can support this conclusion. The usability results revealed that IM-IO, which used icons in both Full Mini-Map and On-Object Indicators, achieved the "Excellent" adjective rating. In the subjective mental workload results (Figure 12), IM-IO outperformed five of the six dimensions: physical demand, temporal demand, performance, effort, and frustration. And it also achieved a relatively low mental demand load. In the participants' feedback, G06 said: *"Iconic shows me a lot which allows me to identity easier and quicker."* G02 added: *"It [the iconic representative method] is something to get use for a bit, but it is easy to understand."*

In terms of using descriptive text as Sound-Type Indicators in Study 2, TM-TO achieved the lowest mental demand in the subjective mental workload scale and a "Good" adjective system usability rating. The participants' feedback supported that the text representation method is easier to understand in most cases. For example, G02 said: *"Texts on the environment work well as it describes the sounds well good enough like phones ringing or waterfall." "[Text has] Nothing lost in translation."* G14 added. However, when it came to a situation where the descriptive text in the Sound-Type Indicators did not exactly match the instruction, the text Sound-Type Indicator visualization may cause a higher cognitive load. For example, the instruction of task No.7 is "Find the object making the campfire sound" (Table 4), while the sound type label of the correct sound source shows "Crackling Fire" (Figure 4c). G16 noted: *"Inconsistency in the wording of objects (prior to task vs during task) made it more difficult to identify the appropriate object."*

5.6.2 Sound Localization. Unlike Study 1, in which the sound sources were evenly organized around the player in a circle, we randomly distributed the sound sources in the VR environment in Study 2 to mimic the sound design of an actual VR game scene. There were sound sources that stayed close to each other, and a sound source was put at position right above another sound source (same azimuth angle, but different altitude). Here, although the Mini-Map Indicator only indicated the directions of the sound sources in a 2D plane and did not show the altitude information of the sound source in a 3D space, we believe that the On-Object Indicator overcame this issue. Unfortunately however, no comment on this aspect was received from the participants.

In Study 2, some tasks had multiple (at most three) sounds happening simultaneously. The tasks of Study 2 with multiple sounding objects were designed to explore if the DHH participants could localize the correct sound source among many with the help of the visualization methods. Based on the analysis of the qualitative results, most of the participants (11/14) indicated that the visualization methods worked well in assisting them to identify the sound source when multiple sounds were happening in the scene. For example, when talking about the TM-TO, which used text to indicate sound types, G10 said: *"It [TM-TO] works well in cases where there are multiple different types of noise and we need to identify a specific sound. This is dependent on how descriptive the text is however."* However, participants also mentioned when the simultaneous sounds contained short duration sounds, they would still struggle to localize the correct sound source, even with the assistance of the visualization methods. For example, G05 noted: *"Works well when there are various sounds going on. Doesn't work well when it quickly flashed and went away when I was trying to read multiple texts to identify the object making the noise."*

5.6.3 Preferences. In terms of the procedure of searching for the sound source using the visualization methods, most of the participants (10/14) indicated that they first used the Full Mini-Map to identify the direction of the correct sound clip, and then use the On-Object Indicator to locate the specific sound source. For example, G06 said: *"I looked at the map where the icon is, then I search around the room to find the source of the sound where the icon marked on the map."* This primarily followed the similar pattern of locating the sound source using FM-OI as indicated by the participants in Study 1.

In addition, we investigated the participants' preferences on the four tested Sound-Type Indicator visualization methods, especially preferences on using text or icons to indicate the sound types. For the overall preferred sound visualization method of Study 2, the participants provided diverse feedback - IM-IO: 4/14, IM-TO: 3/14, TM-IO: 3/14, TM-TO: 4/14. The participants who preferred icons indicated that icons had lower cognitive load, allowing them to focus on the game content. For example, G16 mentioned: *"I thought that iconic indicators are easier/quicker to understand than texts. I do not prefer reading a lot when I play VR games."* The participants who tended to use text as Sound-Type Indicators voiced that text is more noticeable in the scene and can clearly convey its meaning. As noted by G10, *"Text is easier for me to notice on the map, and also to identify the specific type of noise."*

While we could not come to a consensus on what might be the most preferred method for indicating sound types, this also indicates that the future designs of Sound-Type Indicators may allow the user to customize what to be displayed based on their own preferences.

6 DISCUSSION, LIMITATIONS AND FUTURE WORK

The results of Study 1 show that FM-OI had the best performance in localization accuracy and task completion time. The results from the questionnaires reveal that FM-OI has a good system usability and less subjective mental workloads. Also, the participants' feedback shows that FM-OI can allow them to pay more attention to the game content while spending less effort in locating the sound source.

The results of Study 2 show that the four tested Sound-Type Indicators have similar performance on localization accuracy and task completion time. The participants can identify the sound type of a sound source and complete the tasks with the assistance of the tested sound visualization methods. Moreover, the participants have diverse preferences on the four tested Sound-Type Indicators in terms of using text or icons to indicate the sound types in the visualization methods, thus indicating a customizability of the Sound-Type Indicators based on the individual preferences.

There are several limitations we wish to address.

6.1 Limitations

Our current study addressed four sound categories from the sound taxonomy in VR proposed by Dhruv Jain et al. [18], while another five sound categories were not explored here. These unaddressed categories were Non-localized speech, Notification sounds, Interaction sounds, Surrounding ambience, and Music. To provide full immersive and accessible sound experience, our future studies will explore on these unaddressed categories.

We evaluated our sound visualization methods in a scenario with at most three synchronous sound sources in our Study 2. The performance of the sound visualization method in scenarios with more than three concurrent sound sources (e.g., a large crowd of talking people) was untested. In that case, filtering strategies of sounding objects might need to be investigated.

The Full Mini-Map component in our proposed sound visualization methods indicated sounds within its circular range. For the sounds that were outside the range, DHH users can only identify those, that were within their field of view, with the assistance of On-Object Sound Indicators. We aim to explore visualizing the distant sounds on the Full Mini-Map in our future study.

In our second study, as a starting point, we selected the sound type icons and text descriptions (and their display parameters such as the size, black and white color, etc.) based on trial and error with a few pilot studies. However, such content can be changed based on user preferences such as the icon types, shorter/longer descriptions, text and icon sizes/colors, etc. Therefore, these parameters should be investigated in the future from a user preference perspective as well as a content-designer's perspective.

6.2 Future Works

In our future studies, we aim to explore integrating SoundVizVR into different genres of VR applications, especially VR games, to further evaluate its performance. In addition, we aim to explore SoundVizVR in the 3D space [39], how it will affect the immersive experience and also explore other visualization techniques. Options that enable customization to the integrated SoundVizVR components (e.g., changing the position, size, opacity, or background of the Mini-Maps, or changing the size or color of the icon and text in the Sound-Type Indicator) will also be explored in our future studies.

We also intend to explore using the SoundVizVR plugin with VR content designers. This would enable us to determine any preferences of the method and the plugin from a game designer's perspective such as the how easy it is to add this plugin to the workflow, how customizable the experiences are, etc.

In addition, to have a better understanding of SoundVizVR, we are looking forward to applying eye-tracking devices in our user experiments to see how participants perceive information from the user interface. Also, we are looking forward to evaluating SoundVizVR in other user groups (e.g., hearing people) to explore if it could address other accessibility issues, such as situational impairments among the hearing people during their using VR applications [37]. This knowledge may help further improve our designs.

7 CONCLUSION

In this paper, we proposed SoundVizVR that aimed to advance sound accessibility in VR environments for DHH users. We conducted a user experiment with 11 DHH participants to identify the best performing and most preferred Sound-Characteristic Indicator method from six design combinations. Furthermore, to evaluate the performance and user experience of four different Sound-Type Indicator visualization methods, we conducted another user study with 14 DHH participants. Participants' task performance and feedback indicated that SoundVizVR can assist them in locating sound sources, identifying sound characteristics, and identifying the sound types of in-game sound effects in VR environments.

REFERENCES

[1] Belén Agulló, Mario Montagud, and Isaac Fraile. 2019. Making interaction with virtual reality accessible: rendering and guiding methods for subtitles. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 33 (11 2019), 1–13. https://doi.org/10.1017/S0890060419000362

[2] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *J. Usability Studies* 4, 3 (may 2009), 114–123.

[3] Fl'vio Coutinho, Raquel O. Prates, and Luiz Chaimowicz. 2011. An Analysis of Information Conveyed through Audio in an FPS Game and Its Impact on Deaf Players Experience. In *2011 Brazilian Symposium on Games and Digital Entertainment*. 53–62. https://doi.org/10.1109/SBGAMES.2011.16

[4] Sigal Eden and Sara Ingber. 2014. Enhancing Storytelling Ability with Virtual Environment among Deaf and Hard-of-Hearing Children. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Deborah Fels, Dominique Archambault, Petr Peňáz, and Wolfgang Zagler (Eds.). Springer International Publishing, Cham, 386–392.

[5] 3D Everything. 2021. Kids Room | 3D Interior | Unity Asset Store. Retrieved April 13, 2022 from https://assetstore.unity.com/packages/3d/props/interior/kids-room-48596

[6] Facebook. 2022. Oculus Rift S: PC-Powered VR Gaming Headset | Oculus. Retrieved April 13, 2022 from https://www.oculus.com/rift-s/

[7] Epic Games. 2022. Fortnite – a free-to-play Battle Royale game and more. Retrieved April 12, 2022 from https://www.epicgames.com/fortnite/en-US/home

[8] William W. Gaver. 1987. Auditory Icons: Using Sound in Computer Interfaces. *SIGCHI Bull.* 19, 1 (jul 1987), 74. https://doi.org/10.1145/28189.1044809

[9] Kathrin Gerling, Patrick Dickinson, Kieran Hicks, Liam Mason, Adalberto L. Simeone, and Katta Spiel. 2020. *Virtual Reality Games for People Using Wheelchairs.* Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376265

[10] 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* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 599, 14 pages. https://doi.org/10.1145/3411764.3445196

[11] Ru Guo, Yiru Yang, Johnson Kuang, Xue Bin, Dhruv Jain, Steven Goodman, Leah Findlater, and Jon Froehlich. 2020. HoloSound: Combining Speech and Sound Identification for Deaf or Hard of Hearing Users on a Head-Mounted Display. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20)*. Association for Computing Machinery, New York, NY, USA, Article 71, 4 pages. https://doi.org/10.1145/3373625.3418031

[12] S. G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (2006), 904 – 908.

[13] Chris J. Hughes, Marta B. Zapata, Matthew Johnston, and Pilar Orero. 2020. Immersive Captioning: Developing a framework for evaluating user needs. In *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*. 313–318. https://doi.org/10.1109/AIVR50618.2020.00063

[14] 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* (Hong Kong, China) *(DIS '18 Companion)*. Association for Computing Machinery, New York, NY, USA, 7–11. https://doi.org/10.1145/3197391.3205404

[15] Dhruv Jain, Brendon Chiu, Steven Goodman, Chris Schmandt, Leah Findlater, and Jon E. Froehlich. 2020. Field Study of a Tactile Sound Awareness Device for Deaf Users. In *Proceedings of the 2020 International Symposium on Wearable Computers* (Virtual Event, Mexico) *(ISWC '20)*. Association for Computing Machinery, New York, NY, USA, 55–57. https://doi.org/10.1145/3410531.3414291

[16] Dhruv Jain, Leah Findlater, Jamie Gilkeson, Benjamin Holland, Ramani Duraiswami, Dmitry Zotkin, Christian Vogler, and Jon E. 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* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 241–250. https://doi.org/10.1145/2702123.2702393

[17] Dhruv Jain, Khoa Huynh Anh Nguyen, Steven M. Goodman, Rachel Grossman-Kahn, Hung Ngo, Aditya Kusupati, Ruofei Du, Alex Olwal, Leah Findlater, and Jon E. Froehlich. 2022. ProtoSound: A Personalized and Scalable Sound Recognition System for Deaf and Hard-of-Hearing Users. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 305, 16 pages. https://doi.org/10.1145/3491102.3502020

[18] Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John Porter, Chris Yoon, Swetha Machanavajhala, and Meredith Ringel Morris. 2021. A Taxonomy of Sounds in Virtual Reality. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 160–170. https://doi.org/10.1145/3461778.3462106

[19] Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John R. Porter, Chris Yoon, Swetha Machanavajhala, and Meredith Ringel Morris. 2021. Towards Sound Accessibility in Virtual Reality. In *Proceedings of the 2021 International Conference on Multimodal Interaction* (Montréal, QC, Canada) *(ICMI '21)*. Association for Computing Machinery, New York, NY, USA, 80–91. https://doi.org/10.1145/3462244.3479946

[20] 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.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376758

[21] Peter Keller and Catherine Stevens. 2004. Meaning From Environmental Sounds: Types of Signal-Referent Relations and Their Effect on Recognizing Auditory Icons. *Journal of experimental psychology. Applied* 10 (04 2004), 3–12. https://doi.org/10.1037/1076-898X.10.1.3

[22] Firesprite Limited. 2021. The Persistence - Games | PlayStation®. Retrieved April 12, 2022 from https://www.playstation.com/en-us/games/the-persistence/

[23] Carme Mangiron. 2016. Reception of game subtitles: an empirical study. *The Translator* 22, 1 (2016), 72–93. https://doi.org/10.1080/13556509.2015.1110000 arXiv:https://doi.org/10.1080/13556509.2015.1110000

[24] 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 (2006), 333–351.

[25] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[26] Mohammadreza Mirzaei, Peter Kán, and Hannes Kaufmann. 2020. EarVR: Using Ear Haptics in Virtual Reality for Deaf and Hard-of-Hearing People. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 2084–2093. https://doi.org/10.1109/TVCG.2020.2973441

[27] Mojang. 2022. Minecraft Official Site | Minecraft. Retrieved April 13, 2022 from https://www.minecraft.net/en-us

[28] Jorge Mori and Deborah I. Fels. 2009. Seeing the music can animated lyrics provide access to the emotional content in music for people who are deaf or hard of hearing?. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*. 951–956. https://doi.org/10.1109/TIC-STH.2009.5444362

[29] Suranga Nanayakkara, Elizabeth Taylor, Lonce Wyse, and S H. Ong. 2009. An Enhanced Musical Experience for the Deaf: Design and Evaluation of a Music Display and a Haptic Chair. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 337–346. https://doi.org/10.1145/1518701.1518756

[30] Suranga Chandima Nanayakkara, Lonce Wyse, S. H. Ong, and Elizabeth A. Taylor. 2013. Enhancing Musical Experience for the Hearing-Impaired Using Visual and Haptic Displays. *Human–Computer Interaction* 28, 2 (2013), 115–160. https://doi.org/10.1080/07370024.2012.697006 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/07370024.2012.697006

[31] De Xing Ong, Kai Xiang Chia, Yi Yi Huang, Jasper Teck Siong Teo, Jezamine Tan, Melissa Lim, Dongyu Qiu, Xinxing Xia, and Frank Yunqing Guan. 2021. Smart Captions: A Novel Solution for Closed Captioning in Theatre Settings with AR Glasses. In *2021 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. 1–5. https://doi.org/10.1109/SOLI54607.2021.9672391

[32] 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* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 495–499. https://doi.org/10.1145/3308561.3353810

[33] 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 (2001), 681–686.

[34] Yi-Hao Peng, Ming-Wei Hsi, 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 *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3173574.3173867

[35] Sandra Poeschl, Konstantin Wall, and Nicola Doering. 2013. Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence. In *2013 IEEE Virtual Reality (VR)*. 129–130. https://doi.org/10.1109/VR.2013.6549396

[36] Sylvia Rothe, Kim Tran, and Heinrich Hussmann. 2018. Positioning of Subtitles in Cinematic Virtual Reality. https://doi.org/10.2312/egve.20181307

[37] Zhanna Sarsenbayeva, Niels van Berkel, Chu Luo, Vassilis Kostakos, and Jorge Goncalves. 2017. Challenges of Situational Impairments during Interaction with Mobile Devices. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (Brisbane, Queensland, Australia) *(OZCHI '17)*. Association for Computing Machinery, New York, NY, USA, 477–481. https://doi.org/10.1145/3152771.3156161

[38] Stefania Serafin and Giovanni Serafin. 2004. Sound Design to Enhance Presence in Photorealistic Virtual Reality. In *ICAD*.

[39] Lichao Shen, MHD Yamen Saraiji, Kai Kunze, Kouta Minamizawa, and Roshan Lalintha Peiris. 2020. Visuomotor Influence of Attached Robotic Neck Augmentation. In *Symposium on Spatial User Interaction* (Virtual Event, Canada) *(SUI '20)*. Association for Computing Machinery, New York, NY, USA, Article 14, 10 pages. https://doi.org/10.1145/3385959.3418460

[40] 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.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376353

[41] Sean M. Smith and Glen N. Williams. 1997. A Visualization of Music. In *Proceedings of the 8th Conference on Visualization '97* (Phoenix, Arizona, USA) *(VIS '97)*. IEEE Computer Society Press, Washington, DC, USA, 499–ff.

[42] Axel Stockburger. 2003. The game environment from an auditive perspective. *Level Up* (2003), 4–6.

[43] Valday Team. 2016. Vintage 50s Radio and Phone | 3D Electronics | Unity Asset Store. Retrieved April 13, 2022 from https://assetstore.unity.com/packages/3d/props/electronics/vintage-50s-radio-and-phone-71657

[44] Mauro Teófilo, Alvaro Lourenço, Juliana Postal, and Vicente Lucena Jr. 2018. *Exploring Virtual Reality to Enable Deaf or Hard of Hearing Accessibility in Live*

Theaters: A Case Study. 132–148. https://doi.org/10.1007/978-3-319-92052-8_11

[45] vUv. 2021. Speakers PBR | 3D Electronics | Unity Asset Store. Retrieved April 13, 2022 from https://assetstore.unity.com/packages/3d/props/electronics/speakers-pbr-111606

[46] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300371

[47] Minecraft Wiki. 2022. Subtitles – Minecraft Wiki. Retrieved April 12, 2022 from https://minecraft.fandom.com/wiki/Subtitles

[48] Bei Yuan, Eelke Folmer, and Frederick C. Harris. 2011. Game Accessibility: A Survey. *Univers. Access Inf. Soc.* 10, 1 (mar 2011), 81–100. https://doi.org/10.1007/s10209-010-0189-5

[49] Veronica Zammitto. 2008. VISUALIZATION TECHNIQUES IN VIDEO GAMES. https://doi.org/10.14236/ewic/EVA2008.30

[50] 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* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300341