



OPEN ACCESS

EDITED BY

Hai-Ning Liang,
The Hong Kong University of Science and
Technology (Guangzhou), China

REVIEWED BY

Unais Sait,
Free University of Bozen-Bolzano, Italy
Yassine AIDI,
University of Sfax, Tunisia

*CORRESPONDENCE

Takefumi Ogawa,
✉ ogawa@nc.u-tokyo.ac.jp
Arinobu Nijijima,
✉ arinobu.nijijima@ntt.com

RECEIVED 24 March 2025

ACCEPTED 03 July 2025

PUBLISHED 14 July 2025

CITATION

Takahara Y, Nijijima A, Park C and Ogawa T
(2025) Enhancing eyes-free interaction in virtual
reality using sonification for multiple
object selection.
Front. Virtual Real. 6:1598776.
doi: 10.3389/frvir.2025.1598776

COPYRIGHT

© 2025 Takahara, Nijijima, Park and Ogawa. This
is an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Enhancing eyes-free interaction in virtual reality using sonification for multiple object selection

Yota Takahara¹, Arinobu Nijijima^{2*}, Chanh Park¹ and
Takefumi Ogawa^{3*}

¹Graduate School of Engineering, The University of Tokyo, Tokyo, Japan, ²NTT Human Informatics Laboratories, NTT Corporation, Yokosuka, Kanagawa, Japan, ³Information Technology Center, The University of Tokyo, Chiba, Japan

In virtual reality (VR) environments, selecting and manipulating multiple out-of-view objects is often challenging because most current VR systems lack integrated haptics. To address this limitation, we propose a sonification method that guides users' hands to target objects outside their field of view by assigning distinct auditory parameters (pan, frequency, and amplitude) to the three spatial axes. These parameters are discretized into three exponential steps within a comfortable volume (less than 43 dB) and frequency range (150–700 Hz), determined via pilot studies to avoid listener fatigue. Our method dynamically shifts the sound source location depending on the density of the target objects: when objects are sparsely positioned, each object serves as its own sound source, whereas for dense clusters, a single sound source is placed at the cluster's center to prevent overlapping sounds. We validated our technique through user studies involving two VR applications: a shooting game that requires rapid weapon selection and a 3D cube keyboard for text entry. Compared to a no-sound baseline, our sonification significantly improved positional accuracy in eyes-free selection tasks. In the shooting game, participants could more easily swap weapons without losing sight of on-screen action, while in the keyboard task, typing accuracy more than doubled during blind entry. These findings suggest that sonification can substantially enhance eyes-free interaction in VR without relying on haptic or visual cues, thereby offering a promising avenue for more efficient and comfortable VR experiences.

KEYWORDS

sonification, eyes-free interaction, virtual reality, multiple object selection, auditory feedback

1 Introduction

Virtual reality (VR) technologies are used in a variety of applications ranging from gaming (Nor et al., 2019) to medical simulation (Allgaier et al., 2022; Chen et al., 2022), where users must accurately perceive and manipulate three-dimensional (3D) virtual objects (Maslych et al., 2024). Visual feedback is typically essential for confirming the positions of objects and adjusting movements accordingly, thereby improving the precision of operations in VR (Abe et al., 2021). However, not every interaction occurs directly within the user's field of view. Out-of-sight interactions, also known as “eyes-free target acquisition” (Yan et al., 2018), rely on spatial memory and proprioception and can enhance operational efficiency as well as mitigate VR-induced motion sickness. For

instance, in a driving simulator, users must often operate a gear shift located outside their field of view without taking their eyes off the road.

To improve the accuracy of acquiring out-of-view objects in VR, researchers have adopted haptic feedback in selection tasks. Ren et al. (2024) introduced aero-haptic feedback in which users wear a device that directs airflow to their fingertips. The wind originates from the direction of the target, providing a spatial cue, and its intensity increases as the hand approaches the object, supplying a distance cue. Together, these cues help users efficiently locate out-of-view objects. Ariza N. et al. (2017) developed a vibrotactile feedback device that guides users toward out-of-view objects. Their study showed that vibrotactile assistance made object acquisition faster and more accurate, especially when the required rotation angle was small. These studies showed that haptic feedback is effective, but it relies on dedicated hardware.

In practical terms, accurately acquiring out-of-view targets in VR is complicated by the lack of haptic feedback unless that feedback is purposely integrated. Unlike in real-world environments, users cannot rely on touch or tactile sensations to locate objects and confirm selections. Instead, they must rely entirely on spatial memory and proprioception, which leads to lower accuracy compared to tasks performed in the real world. Various methods integrating sensory feedback have been proposed to address this challenge (Gao et al., 2019; Xu et al., 2022). Among these, auditory feedback methods offer strong potential, as they map specific sounds to objects and guide users via audio cues (Gao et al., 2019). Auditory feedback allows several sound parameters to change independently while still being reasonably easy for users to discriminate, and it can be delivered through the built-in speakers found on most VR HMDs.

Despite progress in sound-based techniques, most existing sonification approaches focus on interactions with a single out-of-view target. In other words, although existing approaches use pitch to convey the distance to a single target (Gao et al., 2019) or combine pitch and tempo to indicate its height and direction (Gao et al., 2022), they are not designed to provide auditory feedback that simultaneously represents the positions of multiple objects. Yet in many VR applications, such as typing on a virtual keyboard that users do not look at directly or switching between multiple weapons in a VR shooting game, there can be numerous potential targets (Wu et al., 2023). Users must maintain their visual attention on primary tasks (e.g., observing the main screen or targeting enemies) while interacting with objects located elsewhere. Consequently, a more flexible, multi-target sonification approach is required. In addition, there is still no unified guideline on which auditory parameters to map to which spatial dimensions or how those parameters should vary; design choices remain highly application specific. Mapping sounds for the manipulation of multiple out-of-view objects is virtually unexplored.

In this work, we extend sonification to accommodate multiple out-of-view targets. Our method assigns three distinct auditory parameters (pan, frequency, and amplitude) to the 3D axes, thereby enabling users to discern along which axis and in which direction their hand is misaligned with the target. We further introduce a strategy that adjusts the position of the sound source according to the density of the target objects. When objects are spread out (sparse), the sound source is

placed at each object's center; when objects are packed tightly (dense), the cluster's center becomes the sound source to prevent overlapping audio signals. Through user studies, we demonstrate that our approach enhances the accuracy of eyes-free target acquisition and benefits two practical scenarios: switching weapons in a VR shooting game and performing blind typing on a 3D cube keyboard.

2 Motivation

Our study aims to establish concrete design guidelines validated through user studies for mapping auditory parameters onto 3D space when users manipulate multiple out-of-view objects in VR. When three sound parameters are mapped to the x, y, and z-axes, we ask two research questions:

1. What parameter ranges remain usable in a real VR application?
2. How should those parameters vary so they stay perceptually distinguishable?

Our hypothesis for the first research question is that, given a VR application where the sound itself must not be unpleasant, the sound's frequency and volume should fall within a specific range, one that we expect to be narrower than the full range detectable by human hearing. Our hypothesis for the second research question is that because all three parameters change simultaneously, users will find discrimination harder than when only one parameter varies, resulting in fewer perceptible resolution levels. Because previous work has offered little guidance on how to set these auditory parameters, our study fills this gap by deriving empirically validated guidelines through user studies.

3 Sonification

Sonification is the use of non-speech audio to convey information (Barrass and Kramer, 1999) and has gained recognition for complementing visual displays and supporting kinesthetic perception (Sigrist et al., 2013; Hasegawa et al., 2012). In a typical one-to-one sonification scheme, sound parameters such as pitch, loudness, and panning are mapped to data values, allowing independent and relatively straightforward control of each auditory dimension (Dubus and Bresin, 2013).

Sonification is particularly beneficial when users cannot directly observe their own movements (Geronazzo et al., 2016) or when visual feedback alone is insufficient (Matinfar et al., 2023). For instance, Scholz et al. mapped horizontal mouse movements to pitch and vertical movements to brightness (Scholz et al., 2014). Within extended reality (XR), where virtual elements may be placed in any 3D configuration around a user (Cho et al., 2024), auditory cues can significantly decrease search time for out-of-view targets. One study reported a 35% reduction in search time for a virtual element using auditory guidance (Billinghurst et al., 1998), and sonification in augmented reality (AR) settings has likewise been shown to reduce the time required to locate objects beyond the user's field of view (Binetti et al., 2021). Furthermore, Xu et al. proposed sonification for air-sketching in a mobile AR environment by assigning a French

TABLE 1 Comparison of existing sonification techniques and our method for eyes-free interaction.

Study	#Targets	Controlled parameters
Billinghurst et al. (1998)	Single	Spatialization (3D Pan)
Gao et al. (2019)	Single	Pitch
Binetti et al. (2021)	Single	Spatialization (3D Pan)
Our Method	Multiple	Pan, Frequency, Amplitude Sound Source Position

horn timbre to movements along the x-axis and a cello timbre to the y-axis (Xu et al., 2022).

Designing sonification for multiple dimensions requires careful attention to psychoacoustic principles. Ziemer et al. (Ziemer and Schultheis, 2022) introduced guidelines emphasizing that parameter mapping should be perceptually near-linear (Ziemer and Schultheis, 2020), despite human hearing's tendency to follow a logarithmic scale. Parameters must also exceed the just noticeable difference (JND) so that changes in sound remain clearly discernible (Ziemer and Schultheis, 2019; Barrass and Kramer, 1999), and practitioners should minimize interference between dimensions such as frequency and amplitude, which can affect each other's perception (Schneider, 2018; Zwicker and Fastl, 2013). Moreover, the auditory feedback must remain unobtrusive, so as not to disrupt the primary task or cause prolonged listener fatigue (Pedersen and Sokoler, 1997; Mynatt et al., 1998), underscoring the importance of appropriate sound selection, volume ranges, and parameter resolution.

While researchers have explored single-target acquisition using sonification (Gao et al., 2022; 2019), methods for selecting one object among many remain relatively underdeveloped. Some prior work has combined auditory and tactile cues (Ménélas et al., 2010; May et al., 2019), but few studies have focused on purely auditory approaches to multiple-object selection. Accordingly, we propose a sonification method that encodes both distance and direction to guide users' hands to out-of-view targets without relying on additional haptic devices. By integrating multi-parameter audio cues and strategically placing the sound source depending on target density, our approach offers an effective extension of single-target sonification techniques to multi-target scenarios.

In summary, our sonification technique is novel relative to existing methods because it explicitly addresses the design of sonification for multiple out-of-view selectable objects. Table 1 presents a comparison with previous studies. By increasing the number of control parameters and dynamically relocating the sound source, our approach accommodates scenarios involving several selectable objects.

4 Proposed methods

Our method provides auditory feedback (a “beep” sound) that reflects the positional difference between multiple target objects in a VR environment and the user's hand. By mapping this difference to discrete changes in sound, users can perceive how far and in which direction their hand is offset relative to objects positioned outside

their field of view. This auditory cue enables them to guide their hand more accurately without relying on visual feedback. The key features are as follows:

1. Independent sound parameters for each axis: We assign pan, frequency, and amplitude to the x, y, and z-axes, respectively. Even if two objects are the same distance away, the character of the sound (i.e., left/right panning, pitch, and loudness) will differ according to the axis along which the user's hand is misaligned. This makes it easier to interpret directionality and correct for positional errors.
2. Dynamic sound-source placement: We adjust the reference coordinates of the sound source depending on the spatial density of target objects. When targets are sparsely distributed, the sound source is placed at each object's center. In contrast, for densely clustered objects, we place the sound source at the cluster's center. This strategy prevents confusion that could arise from overlapping audio cues.

4.1 Sound parameter design

We mapped pan, frequency, and amplitude to the x, y, and z-axes by following real-world acoustic cues. Pan naturally corresponds to the left-right (x) dimension, amplitude suits the depth (z) dimension because nearby sounds are louder and distant ones softer, and frequency fits the vertical (y) dimension, echoing prior work (Scholz et al., 2014; Xu et al., 2022) that represents elevation with pitch and is therefore likely to feel intuitive to users.

We vary each parameter according to the relative distance between the user's hand and the designated sound source. Building on our preliminary experiments (Takahara et al., 2024), we use a discrete, exponential mapping for each parameter, as it was found to yield higher accuracy than continuous or linear mappings. We trigger the sound only if the user's hand lies within a certain threshold distance from the sound source (e.g., 6 cm). Within this range, we subdivide the distance into a fixed number of steps (e.g., three steps). Each step changes the corresponding sound parameter in discrete increments (e.g., every 2 cm).

4.1.1 Pan

Along the x-axis, we use panning values from -1 (left) to 1 (right) in discrete steps. By listening for whether the sound is coming from the left, center, or right, users can quickly judge if they need to move their hand along the x-axis and in which direction. The pan is determined as follows:

$$p(i) = p_{\min} + (p_{\max} - p_{\min}) \times \frac{i}{N-1}, \quad i = 0, 1, \dots, N-1,$$

where $p(i)$ is the panning value at step i , $p_{\min}$ is the minimum panning value (e.g., -1), $p_{\max}$ is the maximum panning value (e.g., 1), and N is the total number of discrete steps.

4.1.2 Frequency

For the y-axis, we map distance to frequency. The frequency changes exponentially across steps as follows:

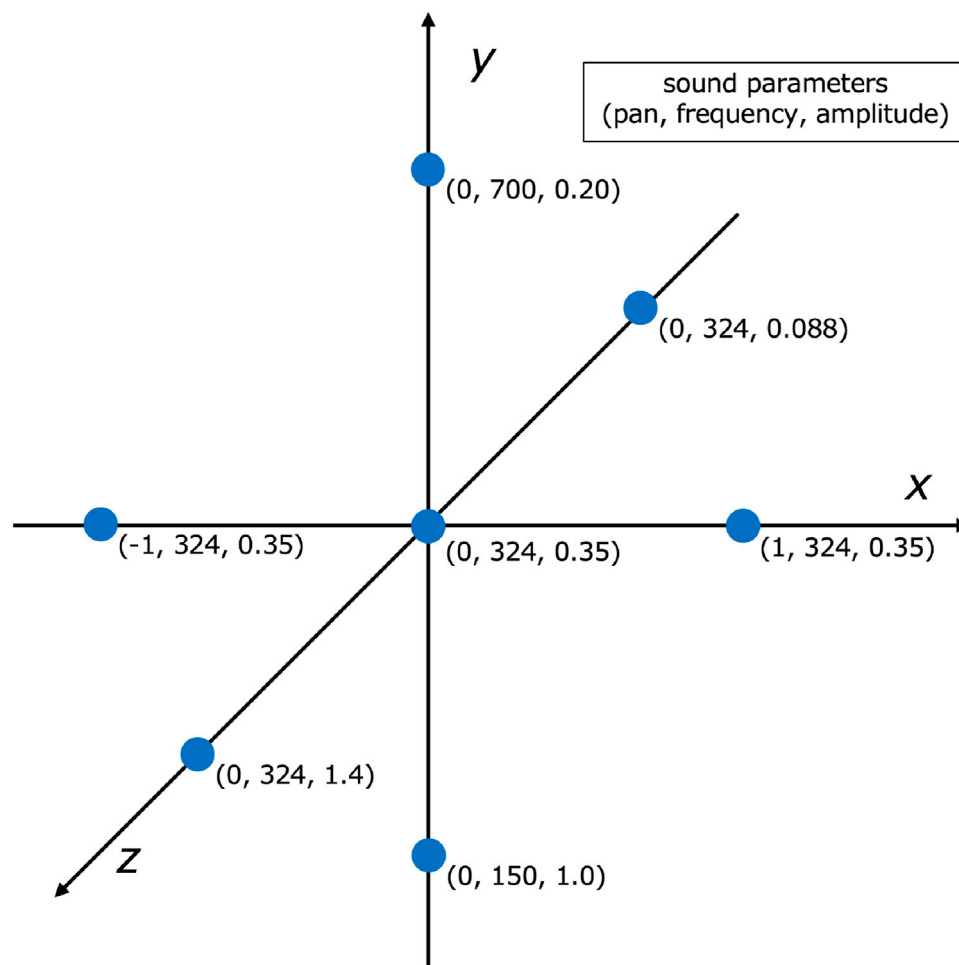


FIGURE 1

An example of how we discretize auditory parameters along each axis. Pan is assigned to the x-axis, frequency to the y-axis, and amplitude to the z-axis. Each parameter is subdivided into three exponential steps, allowing users to discern subtle positional differences without resorting to uncomfortable volume or pitch levels.

$$f(i) = f_{\min} \left(\frac{f_{\max}}{f_{\min}} \right)^{\frac{i}{N-1}}, \quad i = 0, 1, \dots, N-1,$$

where $f(i)$ is the frequency value at step i , $f_{\min}$ is the minimum frequency value, $f_{\max}$ is the maximum frequency value, and N is the total number of discrete steps.

4.1.3 Amplitude

We associate amplitude (loudness) with the z-axis. Like frequency, amplitude is updated in exponential increments across discrete steps, accounting for frequency-dependent loudness perception via correction factors (Ziemer and Schultheis, 2019). This prevents higher pitches from artificially sounding louder than lower pitches at the same amplitude value. The amplitude is determined as follows:

$$a(i) = a_{\min} \left(\frac{a_{\max}}{a_{\min}} \right)^{\frac{i}{N-1}} \times w(f(i)), \quad i = 0, 1, \dots, N-1,$$

where $a(i)$ is the amplitude parameter at step i , $a_{\min}$ is the minimum amplitude value, $a_{\max}$ is the maximum amplitude value, N is the

total number of discrete steps, and $w(f(i))$ is the equal-loudness correction weight for frequency $f(i)$. The weight is calculated based on the ISO 226:2023 Normal Equal-Loudness-Level Contours (Suzuki et al., 2024).

An example of the implementation of our method is shown in Figure 1; each parameter is divided into three discrete steps. The maximum frequency is 700 Hz, and the minimum frequency is 150 Hz. When the frequency is 324 Hz, the maximum amplitude is 1.4, and the minimum amplitude is 0.088. The origin indicates that the hand's position and the sound source's position coincide, and the three axes represent the relative distances between them. Since the sound changes in both the positive and negative directions, the user can determine in which direction the hand is offset relative to the sound source.

4.2 Adjusting sound-source coordinates based on density

Two configurations determine how we position the sound source in relation to multiple objects:

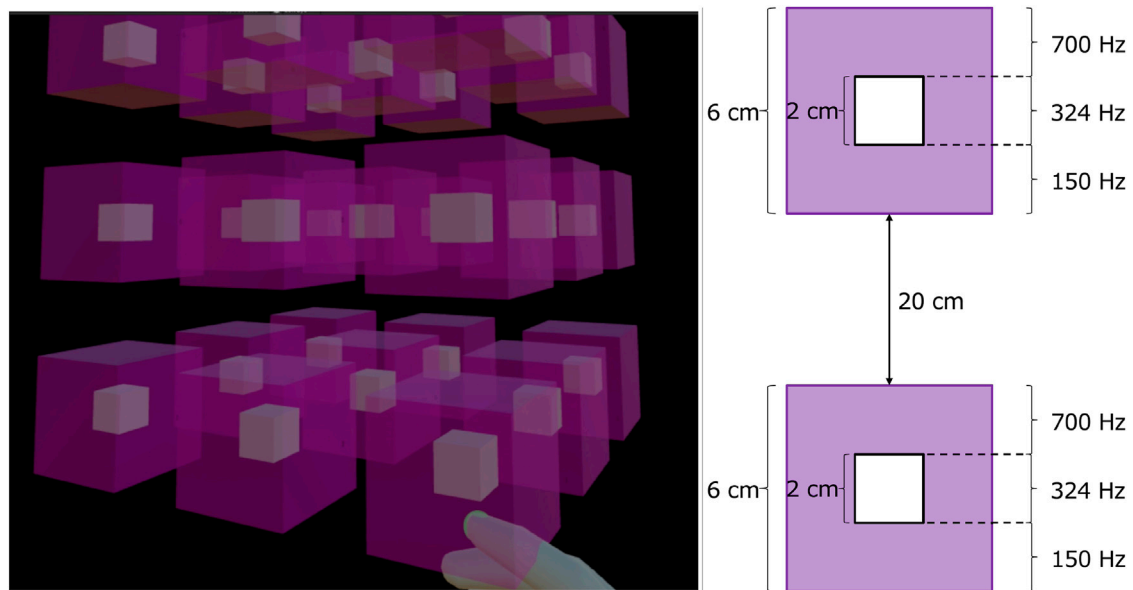


FIGURE 2

When target objects are sparsely arranged, the sound source is positioned at the center of each object. (Left) The purple-shaded areas indicate where auditory feedback is provided. (Right) An example illustrating how the frequency changes in this configuration. The objects are spaced 20 cm apart, and a 6 cm square sound zone is set up around each object where the frequency changes in 2 cm increments. When the distance between the hand and an object is 3 cm or less, a sound is produced, and the pitch of that sound indicates the direction in which the hand is offset relative to the object.

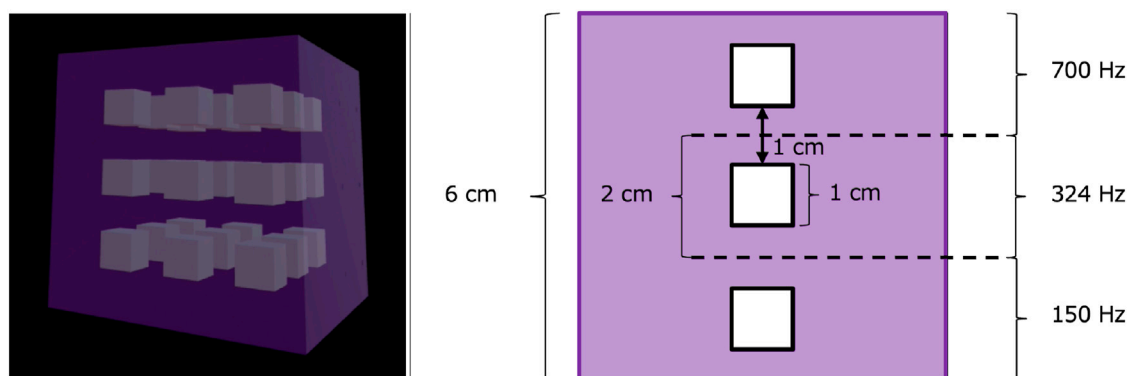


FIGURE 3

When target objects are densely arranged, the sound source is positioned at the center of the object cluster. (Left) The purple-shaded areas indicate where auditory feedback is provided. (Right) An example illustrating how the frequency changes in this configuration is shown. The objects are spaced 2 cm apart, and a 6 cm square sound zone is established that encompasses multiple objects, with the frequency changing in 2 cm increments. When the distance between the hand and the center of the cluster is 3 cm or less, a sound is produced, and its pitch indicates the corresponding object.

1. Sparse distribution (Figure 2): If objects are spaced far apart (e.g., 20 cm or more), we place a sound source at each object's center. We then confine the audible range (e.g., a 6 cm cube around each object), ensuring the sound is only triggered when the user's hand is sufficiently close, thereby preventing unnecessary audio cues.
2. Dense distribution (Figure 3): If objects form a tight cluster (e.g., 4 cm spacing), we place a single sound source at the geometric center of the cluster. The audible zone is expanded slightly to include all objects in that cluster. This approach prevents multiple overlapping audio cues.

In summary, our method capitalizes on discrete, multidimensional auditory cues to guide users' hands toward out-of-view objects in VR. The following sections detail our pilot studies to determine suitable sound parameters and user studies confirming the effectiveness of our approach in VR applications.

5 Applications

We implemented two example applications to demonstrate the feasibility and benefits of our sonification method for out-of-view



FIGURE 4
Maze Chaser Shooter. Players pursue an enemy fleeing through a maze, defeating them by shooting while switching among three types of guns. They select the appropriate weapon from a weapon panel positioned near their right hand, outside their field of view.

interactions in VR: Maze Chaser Shooter and 3D Cube VR Keyboard. Maze Chaser Shooter was developed as a multi-task example in which players navigate a maze, monitor enemy status, select weapons, and take aim. 3D Cube VR Keyboard was created as a single-task example in which users merely type the given character string. Both applications involve effective out-of-view interactions, making them suitable for evaluating the usefulness of our proposed method. Details of evaluation are described in the User Study section.

5.1 Maze Chaser Shooter

In Maze Chaser Shooter (Figure 4), players chase enemies fleeing within a maze, defeating them by shooting while switching among three weapon types. Enemy colors change periodically, and players have to select a matching weapon color to inflict damage. Thus, players need to maintain visual focus on the enemy while selecting the appropriate weapon from a weapon panel positioned outside their field of view on the right side. The weapon panel consists of three vertically arranged cubes (each cube measuring 7.5 cm per side) aligned along the vertical (y -axis). The Dense method is used for sonification, with the sound source placed at the center of the weapon panel cluster. Sounds are triggered when the player's hand is within 11.25 cm of this central sound source. Within this cubic region of 22.5 cm per side, each parameter is set at three levels, changing every 7.5 cm. Thus, by discerning among three distinct sound frequencies, players can accurately switch to the appropriate weapon. Players use the left controller to move their character and the right controller to switch weapons, aim, and shoot.

5.2 3D cube VR keyboard

The second application is a 3D keyboard composed of 27 cubes (each with a side length of 4 cm, arranged in a $3 \times 3 \times 3$ grid), each representing a different alphabet letter (Figure 5). The Dense method is applied for sonification, activating sounds when the user's

hand is within 6 cm of the central sound source. Within a cubic region of 12 cm per side, each auditory parameter is set to change in three distinct levels every 4 cm. Thus, by distinguishing among 27 unique sounds generated by combining three levels of each parameter, users can perform blind typing of all 26 alphabet keys. To facilitate blind typing, users keep their gaze on the main VR display rather than looking down at the cubes. Sonification allows them to pinpoint each letter's location via auditory cues, effectively emulating the tactile feedback missing in VR. By mapping three discrete sound parameters (pan, frequency, and amplitude) to the x -, y -, and z -axes, our method supports accurate eyes-free typing. This setup illustrates how auditory feedback can enhance text-entry tasks in immersive environments where visual attention is directed elsewhere.

6 Pilot studies

Before implementing our sonification method in VR applications, we ran pilot studies to determine suitable sound parameter ranges (amplitude and frequency) and the optimal number of discrete steps per parameter. Although expanding the total number of steps could theoretically convey more information, excessively large sound ranges and overly fine step intervals risk causing discomfort or reduced perceptual accuracy, particularly in extended VR sessions. Our pilot studies, therefore, aimed to identify the largest safe range and highest number of easily distinguishable steps without causing listener fatigue.

All pilot studies were approved by the Ethics Committee of the University of Tokyo (Approval Number: 24-061). Written informed consent was obtained from all participants. All subsequent studies were covered by the same approval.

6.1 Determining comfortable amplitude and frequency ranges

We recruited eight participants (25 ± 3.5 years old), all right-handed.

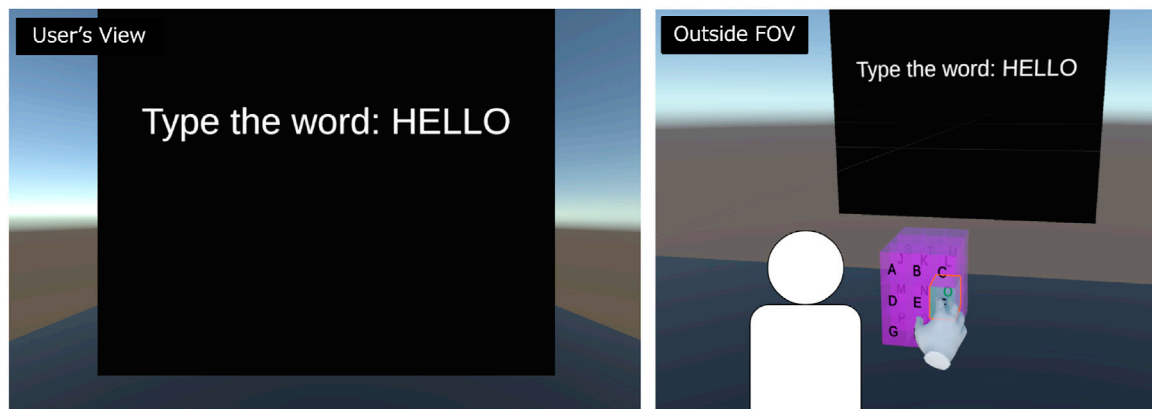


FIGURE 5

3D Cube VR Keyboard. It consists of a 3 × 3 × three arrangement of blocks, each assigned an alphabet letter. By enabling users to input letters outside their field of view, users can perform blind typing while focusing on the input screen, thus improving typing speed.

6.1.1 Procedure

Participants wore headphones (Sony) connected to a Meta Quest 3. We first examined the maximum comfortable amplitude (volume). A sine wave of 440 Hz was played in the center (pan = 0). Using the VR headset's volume controls, participants increased the volume until it became uncomfortable to listen for more than 20 s. The highest volume that remained comfortable was recorded for each participant, and a sound level meter was used to measure its corresponding sound pressure level (dB).

Next, we investigated the comfortable frequency range. A sine wave at 40 dB was used, with its pitch controlled by a slider in Unity. Participants adjusted the slider to indicate the lowest and highest frequencies they could tolerate for at least 20 s without discomfort. These frequencies were recorded for each participant.

6.1.2 Analysis

From the recorded values, we calculated quartiles. For maximum values (sound pressure, frequency), we took the first quartile; for minimum values, we took the third quartile. These quartiles served as global upper or lower boundaries for amplitude and frequency to ensure all participants would find them comfortable during prolonged listening.

6.1.3 Results

We identified 43 dB as the upper bound for comfortable volume and 150–700 Hz as the safe frequency range. Notably, these limits were selected to balance two constraints: (1) the sounds must be sufficiently distinct for reliable cueing, and (2) they must remain non-fatiguing for extended VR sessions.

6.2 Determining the optimal number of discrete steps

Sixteen participants (25 ± 2.4 years old), all right-handed, were recruited.

6.2.1 Procedure

We tested whether participants could reliably distinguish three or five discrete steps in each dimension (pan, frequency,

and amplitude) within the comfortable ranges identified above. A default sound was established at pan = 0, frequency = 324 Hz, amplitude = 0.35. In separate trials, participants focused on one parameter at a time while the other two remained fixed.

Pan values were set to −1, −0.5, 0, 0.5, and 1 for five steps; −1, 0, and 1 for three steps. Frequency values were subdivided exponentially into five or three steps within 150–700 Hz. Amplitude values were subdivided into five or three exponential steps; the loudest possible sound was 1.4 (43 dB), and the softest was about 0.088.

Participants were first given 90 s to familiarize themselves with the sounds. In each single-parameter test, they heard random step values 30 times and responded by indicating which step they thought it was. The number of correct identifications was recorded. The experimental order was varied for each participant to ensure counterbalancing.

They were then tested on all three parameters simultaneously, again conducting three-step and five-step mappings. Each participant completed 30 trials per condition. The number of trials in which all three parameters were correctly identified was recorded.

6.2.2 Results

For single-parameter tests with three steps, the average identification rates were as follows: $99\% \pm 1.4\%$ for pan, $98\% \pm 3.0\%$ for frequency, and $99\% \pm 1.6\%$ for amplitude. For single-parameter tests with five steps, the average identification rates were as follows: $75\% \pm 11\%$ for pan, $83\% \pm 11\%$ for frequency, and $84\% \pm 10\%$ for amplitude. When all three parameters varied simultaneously, overall accuracy was $82\% \pm 10\%$ for three-step mappings, yet fell to $32\% \pm 13\%$ for five-step mappings. These findings suggest that a three-step exponential scheme in each dimension is preferred for practical VR settings. The five-step mapping proved too difficult to discern reliably, especially when users had to process all three parameters at once because each parameter influences the perception of other parameters, as exemplified by phenomena such as frequency-dependent loudness perception.

It should be noted that this experiment was carried out with participants who had no prior training. With extended practice, users might be able to distinguish five levels. For immediate use without considering user training, a three-level distinction appears to be the most practical.

6.3 Summary

Our pilot studies indicate that:

1. Amplitude should not exceed 43 dB, and frequency should remain in the 150–700 Hz range to ensure sustained listening comfort.
2. Three discrete steps per parameter (pan, frequency, amplitude) are optimal for reliable real-time discrimination in a VR environment.

These results informed the design of our sonification system, ensuring the auditory signals are distinct enough to guide eyes-free interactions without causing listener fatigue or confusion.

7 User studies

To confirm the effectiveness of our sonification approach in practical VR contexts, we conducted two user studies using the same two VR applications described earlier: the shooting game (Figure 4) and the 3D keyboard (Figure 5). In the first study, we quantified the positional error in eyes-free target acquisition under different sound-source configurations (sparse vs. dense) and object spacing conditions that mirror our two applications. The second study then integrated our sonification method into the shooting game and 3D keyboard to evaluate user experience and operability, comparing conditions with and without sonification.

7.1 User study 1: error of eyes-free target acquisition

We recruited 16 participants (25 ± 2.4 years old), all right-handed.

7.1.1 Procedure

A total of 27 cubes, each 2 cm on a side, were arranged in a $3 \times 3 \times 3$ grid and placed 20 cm to the right of the participant's head, outside their field of view. Two different center-to-center spacing conditions were tested: (1) 10 cm, approximating the shooting game's weapon panel, and (2) 4 cm, modeling the 3D keyboard layout.

We compared two sonification strategies (Dense and Sparse) against a No-Sound baseline:

1. Dense method: A single sound source is located at the center of the entire cluster.
2. Sparse method: Each of the 27 cubes serves as a sound source.
3. No Sound: No auditory feedback.

In the Dense condition, participants only heard the guiding beep when their hand entered a cubic region surrounding the designated center of the cluster (e.g., 26 cm, i.e., 13 cm around the center, side length for the 10 cm-spaced cubes, 14 cm, i.e., 7 cm around the center, for the 4 cm-spaced cubes), subdivided into three exponential steps per dimension (pan, frequency, amplitude). In the Sparse condition, each cube was surrounded by a smaller 6 cm region, triggering discrete changes in the same three parameters every 2 cm. In the No Sound condition, users relied solely on proprioception.

Participants wore a Meta Quest three headset and used its controller to touch and press the center of a randomly indicated cube. The distance between the cube's center and the press location was recorded as the positional error. Initially, participants completed a 2-min training session to familiarize themselves with the positions of cubes and their corresponding sounds. Afterward, they faced forward and began the experiment without visual access to the cubes. Each participant performed 30 trials per condition, receiving a 2-min break between conditions. The order of the three conditions was counterbalanced to mitigate ordering effects.

7.1.2 Results

Figures 6, 7 summarize the mean positional errors at 10 cm and 4 cm spacing. A Shapiro–Wilk test indicated non-normal distributions, so we used a Friedman test followed by *post hoc* Wilcoxon signed-rank tests with Bonferroni correction.

Under the 10 cm spacing condition, the means and standard deviations for each condition are as follows: 5.8 ± 3.9 cm for the Dense, 7.3 ± 5.6 cm for the Sparse, and 10.9 ± 5.8 cm for the No Sound. In the Sparse condition, the standard deviation was noticeably larger than in the Dense condition. We believe this is because multiple sound sources were present in the Sparse condition. When participants mistakenly reached toward an adjacent, non-target object, the system provided auditory feedback relative to that object's distance, which led them to select the wrong object.

There was a significant difference emerged among the three conditions ($p = 1.1 \times 10^{-5}$). As shown in Figure 6, *post hoc* tests showed that Dense was significantly more accurate than No Sound with a large effect size (Cohen, 1992) ($W = 0.0, adj.p = 9.2 \times 10^{-5}, r = 0.88$). Sparse was not significantly different from Dense ($W = 36.0, adj.p = 0.31, r = 0.41$) or No Sound ($W = 25.0, adj.p = 0.075, r = 0.56$) though the effect sizes were medium to large.

Under the 4 cm spacing condition, the means and standard deviations for each condition are as follows: 2.6 ± 1.5 cm for the Dense, 4.2 ± 2.5 cm for the Sparse, and 8.2 ± 4.2 cm for the No Sound. As in the 10 cm spacing condition, the Dense condition also produced a smaller mean and standard deviation in the 4 cm spacing condition.

There was a significant overall difference ($p = 9.7 \times 10^{-4}$). As shown in Figure 7, *post hoc* comparisons indicated that Dense outperformed both Sparse ($W = 1.0, adj.p = 1.8 \times 10^{-4}, r = 0.41$) and No Sound ($W = 3.0, adj.p = 4.6 \times 10^{-4}, r = 0.84$), while Sparse also surpassed No Sound ($W = 1.0, adj.p = 1.8 \times 10^{-4}, r = 0.56$); the effect sizes were medium to large.

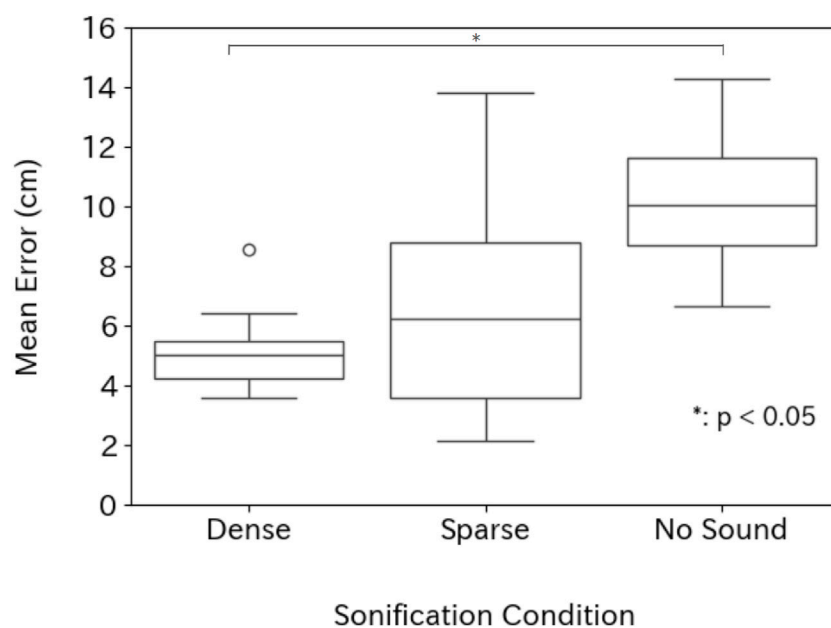


FIGURE 6
Average positional errors for eyes-free target acquisition when adjacent cubes are spaced 10 cm apart.

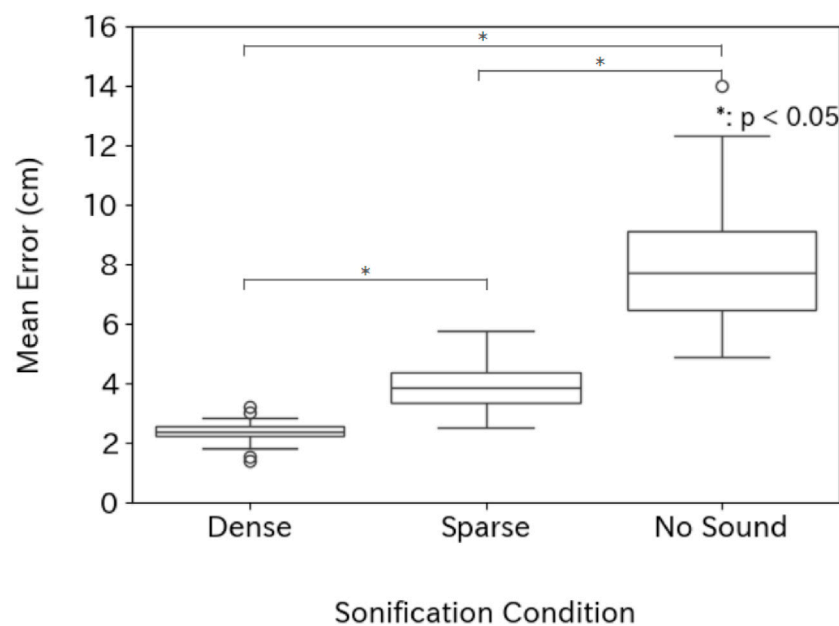


FIGURE 7
Average positional errors for eyes-free target acquisition when adjacent cubes are spaced 4 cm apart.

Overall, the Dense configuration consistently provided the highest accuracy, indicating that placing a single sound source at the cluster's center worked better under both tested spacings. We, therefore, adopted the Dense method for the subsequent user study in our shooting game and 3D keyboard. Note that whether the Dense method or Sparse method is more appropriate depends on the specific conditions. For example, in our previous research (Takahara

et al., 2024), we confirmed that the Sparse method is effective at larger cube spacing, such as 20 cm, compared to that tested in the current experiment. This is because, when objects are spaced about 20 cm apart, the likelihood of unintentionally reaching toward a neighboring object is low, so even in the Sparse condition the auditory feedback still reflects the distance to the correct target object.

7.2 User study 2: Evaluation in VR applications

We recruited 10 participants (25 ± 2.7 years old), all right-handed.

7.2.1 Procedure

Participants wore a Meta Quest three headset and played two VR applications: Maze Chaser Shooter (Figure 4) and 3D Cube VR Keyboard (Figure 5).

In Maze Chaser Shooter, while navigating the maze, participants selected the weapon most effective against the enemy from three weapon panels located outside the field of view and then shoots the enemy. The weapon could be chosen by discriminating among three frequency levels: 150, 324, and 700 Hz. Participants used the left controller to move their character and the right controller to switch weapons, aim, and shoot. Each gameplay session lasted approximately 2–3 min. After each session, participants completed a questionnaire (details described later). Two experimental conditions were tested: with sonification (Sound condition) and without sonification (No Sound condition). Note that gameplay was conducted while background music (BGM) was playing.

In 3D Cube VR Keyboard, participants performed text input using the 3D keyboard composed of 27 cubes arranged in a $3 \times 3 \times 3$ configuration; each cube corresponded to a single alphabet character. Because the auditory parameter associated with each axis changes in three discrete levels, participants could distinguish among 27 unique patterns, enabling them to type the alphabet without looking at the keyboard. The task consisted of a 2-min typing test in which participants performed blind typing of words (e.g., “hello” and “world”) displayed in front of them. Input was performed using the right-hand controller, and any incorrect entries were deleted using the left-hand controller. The experiment was conducted under two conditions: with and without sonification. The accuracy of typed keys was measured under both conditions. Participants completed a questionnaire immediately after the 2-min typing task.

7.2.2 Questionnaire

We adapted items from NASA-TLX (Hart and Staveland, 1988) and prior eyes-free interaction studies (Ren et al., 2024), measuring:

1. Dizziness: Did you feel any VR-induced discomfort?
2. Acquisition Difficulty: How challenging was it to select targets out of view?
3. Distinction Difficulty: How hard was it to differentiate among multiple objects?
4. Mental Strain: How mentally demanding was the task?
5. Fun Experience: How enjoyable was the experience?

All responses were on a 7-point Likert scale.

7.2.3 Results

Figure 8 shows questionnaire responses when playing Maze Chaser Shooter. For the Dizziness, the average scores in both conditions were low, 2.0 in the Sound condition and 2.4 in the No Sound condition; participants did not feel discomfort during the gameplay. For the Acquisition and Distinction Difficulty, the average scores were 2.7 and 2.5 in the Sound condition and those

were 5.1 and 4.8 in the No Sound condition; they could select the appropriate weapon more easily when sonification was presented. For the Mental Strain, the average scores in both conditions were low, 2.5 in the Sound condition and 2.8 in the No Sound condition. For the Fun Experience, the average scores in both conditions were high, 5.4 in the Sound condition and 4.7 in the No Sound condition.

Wilcoxon signed-rank tests with Bonferroni correction revealed significant differences with large effect sizes in Acquisition Difficulty ($W = 0.0, adj.p = 0.010, r = 0.87$) and Distinction Difficulty ($W = 0.0, adj.p = 0.010, r = 0.87$) between Sound and No Sound. No other measures differed significantly. Participants reported that they could more easily locate and switch weapons without looking away from the enemy, confirming the utility of sonification for rapid, eyes-free control. It should be noted that the experiment was conducted under conditions with BGM. In other words, even while BGM was playing, participants could still distinguish the sonification cues and receive effective support for out-of-view operations. This suggests that the sonification could be applied to a wide range of VR applications that include BGM.

Figure 9 shows the questionnaire data when playing 3D Cube VR Keyboard. For the Dizziness, the average scores were 1.5 in both conditions. For the Acquisition and Distinction Difficulty, the average scores were 3.3 and 3.5 in the Sound condition and those were 4.3 and 5.3 in the No Sound condition; Sonification made it easier to recognize the position of each key. For the Mental Strain, the average scores were 3.6 in the both conditions. For the Fun Experience, the average scores were 3.8 in the Sound condition and 3.6 in the No Sound condition. Therefore, the large difference in scores between the two conditions stemmed from how easily the key positions could be recognized.

Distinction Difficulty was significantly lower with a large effect size for Sound compared to No Sound ($W = 0.0098, adj.p = 0.049, r = 0.79$), with other measures showing no significant difference. Average typing accuracy reached 73.2% with sonification, more than double the 29.3% observed without it. The improvement suggests that auditory guidance helps users quickly locate the correct cube while keeping their gaze on the visual display. Note that we did not measure typing speed. Because the participants did not receive sufficient training, typing speed was placed outside the scope of this experiment. Just as users cannot instantly master a new device, operating this VR keyboard as smoothly as a physical keyboard is challenging and requires an adequate period of practice.

When comparing the Maze Chaser Shooter and the 3D Cube VR Keyboard under the Sound condition, participants reported greater difficulty and mental strain with the latter single-task application than with the former multitask game. This appears to stem more from the challenges of out-of-view manipulation than from the overall task complexity. In the Maze Chaser Shooter, users need only choose one weapon from three, whereas in the 3D Cube VR Keyboard they must select one letter from 26, making the task harder and thus more mentally demanding.

7.3 Summary

Our user studies confirm that sonification substantially improves the accuracy and perceived ease of out-of-view

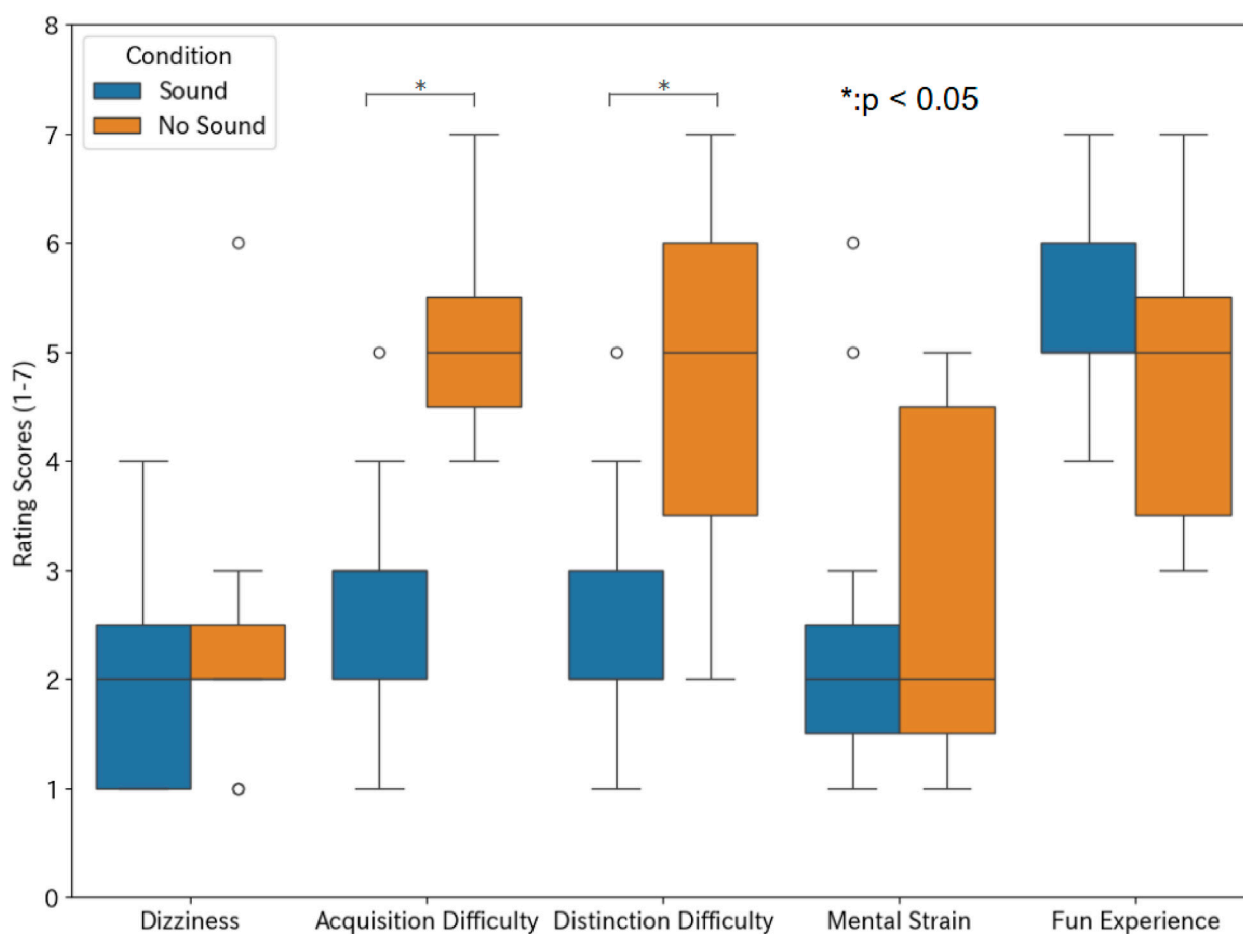


FIGURE 8
Questionnaire results for Maze Chaser Shooter.

interactions in VR. The Dense method reduced positional errors in a multi-target scenario, and VR application tasks like rapidly switching weapons or blind typing on a 3D keyboard became more feasible and less error-prone under auditory guidance. Interestingly, while subjective enjoyment did not significantly differ across conditions, participants often commented that the sound cues gave them greater confidence in their hand positioning, even if they still needed more practice to fully optimize input speed. In conclusion, these findings illustrate the value of carefully designed sonification within comfortable auditory ranges and discrete parameter steps for multi-object selection tasks in VR.

8 Discussion

The primary contribution of this study is the design and validation of a sonification method that facilitates selecting multiple out-of-view objects in VR. Unlike prior sonification approaches that typically target a single, isolated object (Gao et al., 2022; 2019), our technique integrates multi-parameter audio cues and dynamically adjusts the sound source based on the density of target objects. We demonstrated its effectiveness in

two distinct VR applications, a shooting game requiring weapon switching and a 3D keyboard for blind typing, thereby underscoring the method's versatility.

8.1 Key findings

8.1.1 Discrete sonification within comfortable ranges

Our pilot studies established 43 dB and 150–700 Hz as practical, non-fatiguing thresholds for long-term use in VR. By employing discrete exponential steps for pan, frequency, and amplitude, participants could reliably distinguish three levels in each dimension. This balance between comfort and discriminability allowed us to deliver clear audio cues without resorting to overly loud or high-pitched sounds.

8.1.2 Dense vs. sparse sound-source placement

Results from our eyes-free target acquisition tasks showed that a Dense method, using a single sound source at the center of multiple objects, often leads to better accuracy than a Sparse method, at least for object spacings of 4 cm and 10 cm. We surmise that having multiple, closely spaced sources may cause overlap or confusion

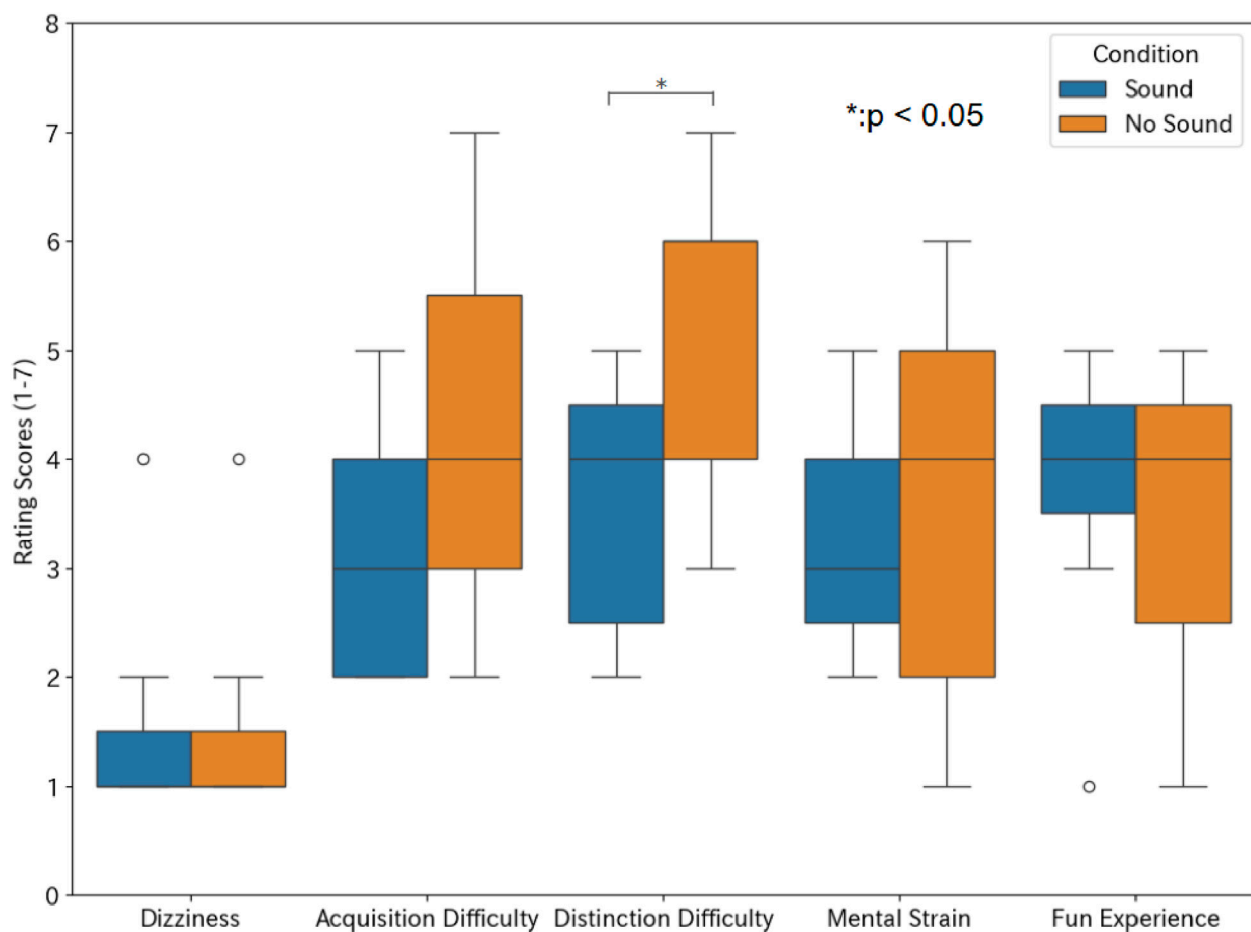


FIGURE 9
Questionnaire results for 3D Cube VR Keyboard.

when objects lie close together. Conversely, a single, central source minimized abrupt shifts in sound parameters. Notably, if objects were spaced very far apart (e.g., 20 cm or more), the Sparse method might again be preferable (Takahara et al., 2024). A more systematic investigation of the thresholds separating “dense” from “sparse” remains an avenue for future work.

In the future, we plan to develop a system that can dynamically adjust sound-source placement. Specifically, when a designer builds a VR application, the system should automatically decide whether to employ the Sparse or Dense method based on the distance between selectable objects. To achieve this, we must run experiments to determine the distance at which the method should switch. Current results indicate that the threshold likely falls between 10 cm and 20 cm.

8.1.3 Practical benefits in VR applications

In Maze Chaser Shooter, participants reported greater ease in locating the correct weapon without diverting their gaze from on-screen action, aligning with the significant improvements in acquisition and distinction difficulty. For the 3D Cube VR Keyboard, typing accuracy increased from 29.3% to 73.2% under sonification, suggesting that auditory cues can effectively replace the tactile feedback that is typically lacking in VR environments.

Our sonification approach is designed for out-of-view interactions and does not require visual input. Consequently, it can be used not only by sighted users but also by individuals with visual impairments. We believe that sonification could effectively assist visually impaired users in perceiving the positions of objects within a VR environment.

8.2 Design guidelines

Our findings emphasize the importance of balancing auditory comfort with sufficient parameter resolution. For prolonged VR sessions, sonification strategies must avoid causing ear fatigue or becoming a source of distraction. By restricting sound pressure levels to approximately 43 dB and frequency changes to the 150–700 Hz band, we ensured the signals were both discernible and tolerable for extended periods. Designing discrete steps (e.g., three exponential levels) also proved essential for accurate identification, especially when pan, frequency, and amplitude were modulated concurrently.

As in the user study conducted for this paper, where users relied on sonification without prior training, it is advisable to limit each auditory parameter to three discrete levels; beyond that, distinguishing the sounds becomes markedly more difficult. Even with just three

levels, the combination of three parameters yields 27 distinct sounds, which our results suggest is more than adequate for applications such as the VR game or typing task used in this experiment.

Although our user study demonstrated the effectiveness of the sonification method even with background music, additional factors must be considered when applying it to VR applications. For example, further investigation is needed into the effects of hand tremor and the limits of spatial memory.

8.3 Limitations and future work

Several limitations in this study suggest directions for future research.

8.3.1 No formal training period

We did not extensively measure how performance or subjective ratings might improve over longer practice. A longitudinal study could shed light on the learning curve associated with multi-object sonification, particularly regarding potential speed gains once users become proficient.

8.3.2 Single-session assessments

Our experiments primarily focused on short-term tasks. Long-term usability and user fatigue (both auditory and cognitive) warrant deeper investigation, especially in scenarios where continuous sonification might be active for extended gaming, simulation, or productivity tasks.

8.3.3 Context-specific optimal parameters

While 43 dB and 150–700 Hz worked well in our applications, other VR setups or user populations (e.g., older adults or people with hearing impairments) may require alternative parameter ranges. Future research could tailor these ranges or step sizes to specific demographics.

8.3.4 Speed of interaction

Although accuracy and subjective ease improved, we did not observe a significant reduction in task completion time. Further studies could evaluate whether more sustained practice allows users to move faster than purely visual methods, verifying the potential of sonification for efficiency gains in eyes-free VR tasks.

9 Conclusion

This paper proposed a sonification technique that makes it easier to manipulate out-of-view objects in VR environments. Unlike previous approaches that focus on selecting a single object, our method establishes design guidelines for assigning sounds when multiple selectable objects are present. Assuming use in VR applications, we imposed the constraint that sound variations must remain comfortable even during prolonged listening and investigated how those sounds should change so they remain easy to distinguish within this limit. Through a series of user studies, we derived the following guidelines:

1. Parameter mapping: Assign pan, frequency, and amplitude to the three spatial axes, varying each parameter according to the positions of the sound source and the user's hand.
2. Parameter ranges: Set frequency to 150–700 Hz and amplitude to less than 43 dB.
3. Discrete levels: Vary each parameter across three levels.
4. Sound-source placement: Determine placement based on object density. Position a sound source at each object's center when objects are sparsely distributed, and at the center of the group when objects are densely packed.

Applying these guidelines to a VR shooting game and a VR keyboard-typing task demonstrated that the sonification strongly supports out-of-view interaction. Because sonification can be delivered solely through a VR HMD's built-in speakers, it is easy to apply our sonification method to various VR applications. Future work will address new challenges such as the effects of practice, long-term use, and expanding the user base to older adults and people with visual impairments.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of the University of Tokyo (Approval Number: 24-061). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

YT: Formal Analysis, Validation, Software, Conceptualization, Methodology, Writing – original draft, Data curation, Investigation, Visualization. AN: Writing – review and editing, Supervision, Project administration, Methodology, Visualization, Conceptualization, Investigation, Validation. CP: Writing – review and editing, Investigation, Supervision. TO: Supervision, Writing – review and editing, Project administration.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

Author AN was employed by NTT Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. The authors declare that Generative AI was used in the creation of this manuscript. We used ChatGPT o1-pro mode (<https://chatgpt.com/>) and DeepL (Version 25.3.31833266, <https://deepl.com/>) to edit and refine written content.

References

- Abe, M., Nakajima, K., and Niitsuma, M. (2021). An investigation of stereoscopic vision and real-scale body motions in a virtual game. *J. Jpn. Soc. Fuzzy Theory Intelligent Inf.* 33, 651–656. doi:10.3156/jsoft.33.3_651
- Allgaier, M., Chheang, V., Saalfeld, P., Apilla, V., Huber, T., Huettl, F., et al. (2022). A comparison of input devices for precise interaction tasks in vr-based surgical planning and training. *Comput. Biol. Med.* 145, 105429. doi:10.1016/j.combiomed.2022.105429
- Ariza, N. O. J., Lange, M., Steinicke, F., and Bruder, G. (2017). "Vibrotactile assistance for user guidance towards selection targets in vr and the cognitive resources involved," in *2017 IEEE symposium on 3D user interfaces (3DUI)*, 95–98. doi:10.1109/3DUI.2017.7893323
- Barras, S., and Kramer, G. (1999). Using sonification. *Multimed. Syst.* 7, 23–31. doi:10.1007/s005300050108
- Billinghurst, M., Bowskill, J., Dyer, N., and Morphet, J. (1998). "An evaluation of wearable information spaces," in *Proceedings. IEEE 1998 virtual reality annual international symposium (cat. No.98CB36180)*, 20–27. doi:10.1109/VRAIS.1998.658418
- Binetti, N., Wu, L., Chen, S., Kruijff, E., Julier, S., and Brumby, D. P. (2021). Using visual and auditory cues to locate out-of-view objects in head-mounted augmented reality. *Displays* 69, 102032. doi:10.1016/j.displa.2021.102032
- Chen, C., Yarmand, M., Singh, V., Sherer, M. V., Murphy, J. D., Zhang, Y., et al. (2022). "Exploring needs and design opportunities for virtual reality-based contour delineations of medical structures," in *Companion of the 2022 ACM SIGCHI symposium on engineering interactive computing systems* (New York, NY, USA: Association for Computing Machinery), 19–25. doi:10.1145/3531706.3536456
- Cho, H., Yan, Y., Todi, K., Parent, M., Smith, M., Jonker, T. R., et al. (2024). "Minexr: mining personalized extended reality interfaces," *Proc. 2024 CHI Conf. Hum. Factors Comput. Syst.* 24, 1–17. doi:10.1145/3613904.3642394
- Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi:10.1037/0033-2909.112.1.155
- Dubus, G., and Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS one* 8, e82491. doi:10.1371/journal.pone.0082491
- Gao, B., Lu, Y., Kim, H., Kim, B., and Long, J. (2019). "Spherical layout with proximity-based multimodal feedback for eyes-free target acquisition in virtual reality," in *HCI 2019* (Springer), 44–58. doi:10.1007/978-3-030-21607-8_4
- Gao, Z., Wang, H., Feng, G., and Lv, H. (2022). Exploring sonification mapping strategies for spatial auditory guidance in immersive virtual environments. *ACM Trans. Appl. Percept.* 19, 1–21. doi:10.1145/3528171
- Geronazzo, M., Bedin, A., Brayda, L., Campus, C., and Avanzini, F. (2016). Interactive spatial sonification for non-visual exploration of virtual maps. *Int. J. Human-Computer Stud.* 85, 4–15. doi:10.1016/j.ijhcs.2015.08.004
- Hart, S. G., and Staveland, L. E. (1988). Development of nasa-tlx (task load index): results of empirical and theoretical research. *Adv. Psychol. (Elsevier)* 52, 139–183. doi:10.1016/S0166-4115(08)62386-9
- Hasegawa, S., Ishijima, S., Kato, F., Mitake, H., and Sato, M. (2012). "Realtime sonification of the center of gravity for skiing," in *Proceedings of the 3rd augmented human international conference* (New York, NY, USA: Association for Computing Machinery), 1–4. doi:10.1145/2160125.2160136
- Maslych, M., Yu, D., Ghasemaghaei, A., Hmaiti, Y., Martinez, E. S., Simon, D., et al. (2024). "From research to practice: survey and taxonomy of object selection in consumer vr applications," in *2024 IEEE international symposium on mixed and augmented reality (ISMAR)*, 990–999. doi:10.1109/ISMAR62088.2024.00115
- Matinfar, S., Salehi, M., Suter, D., Seibold, M., Dehghani, S., Navab, N., et al. (2023). Sonification as a reliable alternative to conventional visual surgical navigation. *Sci. Rep.* 13, 5930. doi:10.1038/s41598-023-32778-z
- May, K. R., Sobel, B., Wilson, J., and Walker, B. N. (2019). Auditory displays to facilitate object targeting in 3d space. *Proc. 2019 Int. Conf. Auditory Disp.* 8, 155–162. doi:10.21785/ICAD2019.008
- Ménelas, B., Picinalli, L., Katz, B. F. G., and Bourdot, P. (2010). "Audio haptic feedbacks for an acquisition task in a multi-target context," in *2010 IEEE symposium on 3D user interfaces (3DUI)*, 51–54. doi:10.1109/3DUI.2010.5444722
- Mynatt, E. D., Back, M., Want, R., Baer, M., and Ellis, J. B. (1998). "Designing audio aura," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.* 98, 566–573. doi:10.1145/274644.274720
- Nor, N. N., Sunar, M. S., and Kapi, A. Y. (2019). A review of gamification in virtual reality (vr) sport. *EAI Endorsed Trans. Creative Technol.* 6, 163212. doi:10.4108/eai.13-7-2018.163212
- Pedersen, E. R., and Sokoler, T. (1997). "Aroma: abstract representation of presence supporting mutual awareness," 97. New York, NY, USA: Association for Computing Machinery, 51–58. doi:10.1145/258549.258584
- Ren, X., He, J., Han, T., Liu, S., Lv, M., and Zhou, R. (2024). Exploring the effect of fingertip aero-haptic feedforward cues in directing eyes-free target acquisition in vr. *Virtual Real. and Intelligent Hardw.* 6, 113–131. doi:10.1016/j.vrih.2023.12.001
- Schneider, A. (2018). *Pitch and pitch perception*. Springer Handbook of Systematic Musicology, 605–685. doi:10.1007/978-3-662-55004-5_31
- Scholz, D. S., Wu, L., Pirzer, J., Schneider, J., Rollnik, J. D., Großbach, M., et al. (2014). Sonification as a possible stroke rehabilitation strategy. *Front. Neurosci.* 8, 332. doi:10.3389/fnins.2014.00332
- Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic Bull. and Rev.* 20, 21–53. doi:10.3758/s13423-012-0333-8
- Suzuki, Y., Takeshima, H., and Kurakata, K. (2024). Revision of iso 226 "normal equal-loudness level contours" from 2003 to 2023 edition: the background and results. *Acoust. Sci. Technol.* 45, 1–8. doi:10.1250/ast.e23.66
- Takahara, Y., Nijima, A., Park, C., and Ogawa, T. (2024). "Minimizing errors in eyes-free target acquisition in virtual reality through auditory feedback," in *Proceedings of the 2024 ACM symposium on spatial user interaction* (New York, NY, USA: Association for Computing Machinery), 24, 1–2. doi:10.1145/3677386.3688884
- Wu, Z., Yu, D., and Goncalves, J. (2023). "Point- and volume-based multi-object acquisition in vr," in *Human-computer interaction – interact 2023* (Springer), 20–42. doi:10.1007/978-3-031-42280-5_2
- Xu, H., Lyu, F., Huang, J., and Tu, H. (2022). Applying sonification to sketching in the air with mobile ar devices. *IEEE Trans. Human-Machine Syst.* 52, 1352–1363. doi:10.1109/THMS.2022.3186592
- Yan, Y., Yu, C., Ma, X., Huang, S., Iqbal, H., and Shi, Y. (2018). "Eyes-free target acquisition in interaction space around the body for virtual reality," in *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–13. doi:10.1145/3173574.3173616
- Ziemer, T., and Schultheis, H. (2019). "Psychoacoustical signal processing for three-dimensional sonification," in *Proceedings of the 25th international Conference on auditory display (ICAD 2019) department of computer and information sciences* (Newcastle upon Tyne, United Kingdom: Northumbria University), 277–284. doi:10.21785/icad2019.018
- Ziemer, T., and Schultheis, H. (2020). Linearity, orthogonality, and resolution of psychoacoustic sonification for multidimensional data. *J. Acoust. Soc. Am.* 148, 2786. doi:10.1121/1.5147752
- Ziemer, T., and Schultheis, H. (2022). Pampas: a psychoacoustical method for the perceptual analysis of multidimensional sonification. *Front. Neurosci.* 16, 930944. doi:10.3389/fnins.2022.930944
- Zwicker, E., and Fastl, H. (2013). "Psychoacoustics: facts and models," 22. Springer Science and Business Media.