New Ears: An Exploratory Study of Audio Interaction Techniques for Performing Search in a Virtual Reality Environment

Muzhe Wu *† Yi Fei Cheng *† David Lindlbauer‡

Human-Computer Interaction Institute
Carnegie Mellon University

2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3315-1647-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISMAR62088.2024.00053

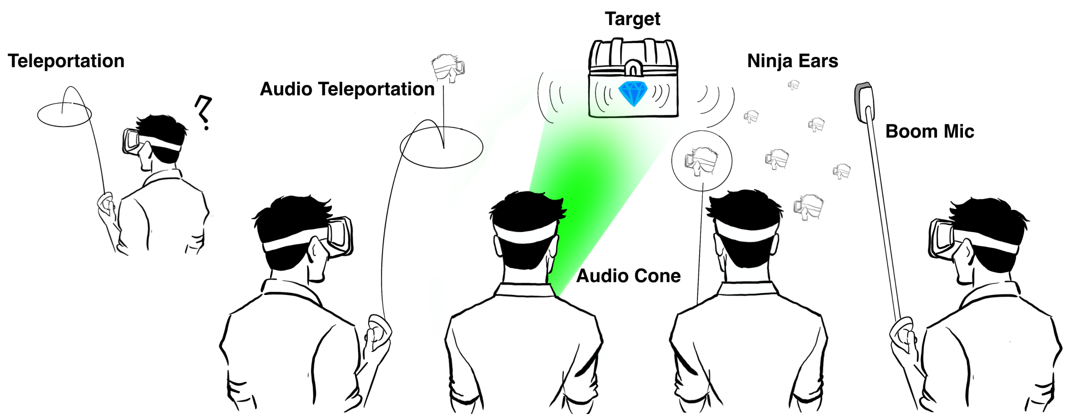

Figure 1: Our 25-participant study evaluated four audio interaction techniques for navigating and searching a virtual environment. *Audio Teleportation* is a direct auditory analogy to *Teleportation*, allowing users to select the target location of their auditory perception. *Audio Cone* invokes a flashlight metaphor, allowing users to project a cone-shaped beam to auditorily amplify sound objects. *Ninja Ears* allows users to select from a grid of potential hearing perspectives. In *Boom Mic*, users hear through a microphone mounted on the end of a boom pole. The study compared these techniques to the conventional *Teleportation* technique and showed effects on search efficiency and user experience, providing implications for future audio interaction technique designs.

Abstract

Efficiently searching and navigating virtual scenes is essential for performing various downstream tasks and ensuring a positive user experience in VR. Prior VR interaction techniques for such scenarios predominantly rely on users' visual perception, which contrasts with physical reality, where people typically rely on multimodal information, especially auditory cues, to guide their spatial awareness. In this work, we explore the potential of leveraging auditory interaction techniques to enhance spatial navigation in virtual environments. We drew inspiration from prior distant interaction techniques and developed four approaches to augmenting how users hear in the virtual environment: Audio Teleportation, Audio Cone, Ninja Ears, and Boom Mic. In a comparative user study ($N = 25$), we evaluated these approaches against a baseline teleportation technique in a search task, where participants traversed a virtual environment to locate target items. Our results suggest that several of our audio interaction techniques may enable more efficient search behaviors while enhancing overall user experience. However, not all techniques were appreciated equally, suggesting that careful attention to their design is critical for ensuring their effectiveness. We conclude by discussing the potential implications of our results for future audio interaction technique designs.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction techniques

*These authors contributed equally to this research.
†e-mail: {muzhew, yifeic2}@andrew.cmu.edu
‡e-mail: davidlindlbauer@cmu.edu

1 Introduction

Search tasks involve scanning an environment and locating target objects [51]. They are performed regularly in everyday life (e.g., relocating one's mobile phone, looking for a book in a library) and are a fundamental interaction in VR experiences. For instance, many VR games involve navigating to a point of interest. VR training applications similarly involve locating task-relevant items to proceed through scenarios. The ability to efficiently search and navigate through virtual environments is thus critical for effective completion of downstream tasks and the general user experience [42].

The need for efficient search and navigation mechanisms has prompted a substantial amount of research on locomotion approaches [2], navigation aids [12], and distant interaction techniques [70]. Many techniques mimic how people navigate in the real world, like *real walking* [69], to offer users an intuitive way to engage with virtual environments. Other techniques embrace the flexibility of the virtual environment, such as *teleportation* [11], to give users capabilities beyond what is possible in the real world (e.g., enabling them to traverse large distances more quickly with lower levels of physical exertion). Currently, most VR interaction techniques rely on the user's *visual* sense to traverse through the environment. In the real world, however, we rely on multiple sensory modalities to engage with our environment.

Our auditory sense is one modality beyond vision that informs our spatial awareness. It enables localization even when vision is limited, such as outside our view or during visual occlusion [62]. Building on these rich perceptual capabilities, many auditory interfaces have been developed for navigation, particularly for blind or low vision users [45]. Auditory cues are also commonly used in virtual environments to inform users of points of interest [3].

979-8-3315-1647-5/24/$31.00 ©2024 IEEE
DOI 10.1109/ISMAR62088.2024.00053

However, in contrast to their visual counterparts, auditory interactions for navigation are primarily grounded in reality, representing a missed opportunity.

In this work, we propose and evaluate four auditory interaction techniques (*Audio Teleportation*, *Audio Cone*, *Ninja Ears*, *Boom Mic*) for navigation and search. They were designed specifically to exploit the flexibility of the virtual environment, offering “beyond real” [1] or “superhuman” [26] hearing capabilities. *Audio Teleportation* builds on conventional teleportation techniques by allowing the user to receive an auditory preview of the target destination before initiating the teleportation. *Audio Cone* invokes the metaphor of a flashlight, serving as a directional audio amplifier to support the user in identifying targets within their environment. Inspired by Ninja Cursor [38] and Ninja Hands [57], *Ninja Ears* offers a selection of auditory perspectives that the user can toggle between to explore their environment. Lastly, *Boom Mic* allows users to control the location from which they hear using a virtual microphone attached to a rigid extension.

We report on the results of a 25-participant empirical study, where we compared *Audio Teleportation*, *Audio Cone*, *Ninja Ears*, and *Boom Mic* to a baseline *Teleportation* technique in a search task. We quantitatively assessed user performance using metrics like task completion time and traversal distance. With questionnaires, we also evaluated participants’ preferences, as well as their perceptions of the techniques’ usability, presence, and embodiment.

Our results show that all techniques reduced traversal distance compared to the baseline. Participants preferred *Audio Teleportation* and *Audio Cone* over *Teleportation* and perceived them as more usable. *Boom Mic*, on the other hand, led to slower performance times and was the least preferred and usable.

In summary, we contribute and evaluate four novel audio interaction techniques for performing search in VR. The techniques notably offer mechanisms for informing the user’s awareness of target locations without requiring them to move. This may beneficially reduce visuo-motor incongruencies, thus reducing motion sickness and discomfort [41], and spatial disorientation [9]. Our results provide insights into the effective design of audio interaction techniques for navigation in VR applications, expanding the space for search tasks beyond the visual modality.

2 Background and Related Work

Our work builds on prior research on navigation in VR, distant interaction techniques, and auditory interactions.

2.1 Navigation in VR

Navigation is a fundamental human activity [17] and a core interaction in VR [42]. It refers to the cognitive and physical processes of goal-directed movement through an environment [48], and has been studied extensively in cognitive science [27] and HCI [34].

In VR, effective interaction techniques for navigation are critical not only because it is in itself a universal task, but necessary as a supporting task in many scenarios, such as reaching check-point locations in tutorials [42]. Navigation techniques should ideally facilitate efficient movement to target locations while minimizing motion sickness and spatial disorientation [6, 68, 54]. Depending on the application, their impact on the user’s sense of presence and embodiment may also be important considerations [65, 18].

Over the years, a variety of VR navigation techniques have been explored [10]. The *walking-in-place* technique, for instance, adapted the walking metaphor for stationary usage [43]. With the *push-button-fly* technique, users steered their direction of movement with their hands [67]. The now widely-implemented *teleportation* technique allows users to point where they want to be in order to shift their viewpoint to that position [11]. All prior techniques have trade-offs regarding their efficiency and impact on motion sickness, spatial orientation, presence, and immersion. For instance, while teleportation offers intuitive and efficient control, it introduces rapid viewpoint changes that can increase spatial disorientation [54]. Despite significant prior work, developing effective VR navigation techniques remains a challenge today.

One limitation of prior VR navigation techniques is that they predominantly focused on the visual experience of navigation, whereas in reality, people navigate relying on multimodal, especially auditory, cues (e.g., [49]). While there is some literature on using audio to facilitate navigation in VR, particularly for accessibility purposes, unlike their visual counterparts, audio-based approaches for navigation are often designed merely to mimic reality (e.g., sonification [23, 47], providing spatialized audio cues [19, 21, 14, 15]), neglecting opportunities in going “beyond-being-real” [1].

In our work, we contribute four audio interaction techniques to facilitate navigation. Our techniques draw inspiration from and build upon prior visual navigation approaches, such as teleportation, and specifically exploit the flexibility of virtual environments to offer *beyond-real* hearing capabilities. In particular, they enable users to audibly explore their environment without moving, which may help them make more informed decisions about their movements and reduce visuo-motor incongruities and sudden viewpoint changes, ultimately supporting a better user experience.

2.2 Distant Interaction Techniques

An alternative to navigating to points of interest in VR is for users to interact with them from a distance. One line of work used abstract interface metaphors for this purpose, like cursors [29], raycasts [31], quad cones [39, 44], and pointers [20]. Among these techniques, our *Audio Cone* interaction draws particular inspiration from the quad cone metaphor, adapting it for the auditory domain.

Also closely related to our work are techniques that leverage an altered representation of the user and their perceptual faculties. Go-Go [53], for instance, introduced a non-linear mapping function between the locations of the physical hand and its virtual representation. Ninja Hands [57] enabled users to control multiple spatially distributed hands to improve target selection. OVRlap [58] let users visually perceive multiple places simultaneously to speed up visual search. However, similar to navigation, distant interactions have thus far primarily focused on the visual modality. One notable exception is HearThere [56], which enables users to hear from spatially distributed locations. Our work builds on this notion of spatially distributed hearing, proposing and evaluating four novel auditory interaction techniques designed to facilitate navigation by augmenting the user’s hearing to offer previews of distant locations.

2.3 Sonic Interactions

There has long been a consensus that audio is an important but underused modality in computing systems [25], including VR [62]. With the emergence of spatial audio technologies, interest in sonic or auditory interactions has grown substantially in recent years [61]. A significant portion of earlier work focused on how audio offers unique affordances for information delivery (e.g., [8, 24, 4, 50, 7]). More recent investigations have greatly expanded the scope of the audio interaction space, incorporating considerations like space, embodiment, semantics, and dynamic contexts (e.g., [30, 52, 14, 33, 22, 72]). From this research, prior work on audio-based navigation is particularly relevant to our work (e.g., [45]).

The human auditory system contains a rich set of perceptual processes for localizing sounds omnidirectionally [49]. It thus serves as a powerful modality for informing navigation [62], especially when the visual sense is impaired. This has prompted a significant amount of prior work to develop and evaluate systems for audio-based navigation. Early work by Loomis et al. [45], for instance, leveraged spatialized audio cues to provide guidance towards target way-points for blind and low vision users. More recent work by Clemenson et al. [17] demonstrated that the use of spatial audio

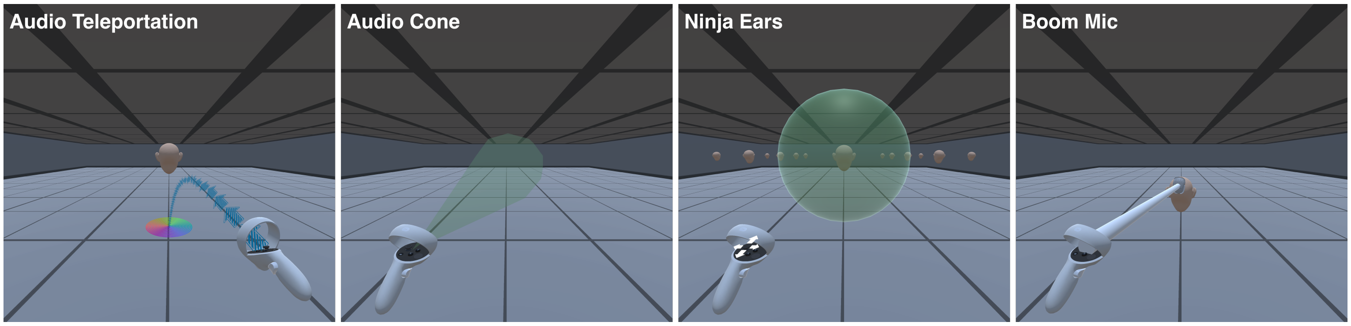

Figure 2: First-person perspectives of our proposed techniques.

for navigation further benefits sighted users by encouraging them to adopt a more active role in their own spatial navigation, leading to more accurate cognitive maps of space.

Overall, prior work on audio-based navigation primarily focuses on informing the user's perception of their surroundings. Our work instead explores how navigation may be enhanced with augmented audio perception faculties, namely the ability to separate one's audio perspective from one's visual viewpoint. We draw inspiration from Geronazzo et al. [26], who prototyped and evaluated three "superhuman" audio beam-forming interaction techniques for enhancing attention. In this work, we similarly prototyped and evaluated four "superhuman" audio interaction techniques aimed at supporting more efficient navigation.

3 New Ears Interaction Techniques

As an initial exploration of auditory interactions for search and navigation, we developed four techniques that extend upon users' auditory perception, shown in Figure 2. We developed these techniques through revisiting classical teleportation and distant-reaching techniques. The four techniques vary across several dimensions, including whether they enabled "superhuman" hearing along a particular direction or at a distant location, whether they were integrated with the locomotion technique, their visual feedback, and the extent to which they were grounded in physical reality. We describe their designs in the following.

3.1 *Audio Teleportation*

As a direct auditory analogy to the *Teleportation* [11], *Audio Teleportation* allows users to select the target location of their auditory perception. *Audio Teleportation* directly integrates with a conventional *Teleportation* technique to facilitate efficient usage.

When the user initiates a teleportation, a visual cue is typically spawned to facilitate selection of a target destination (e.g., [11]). *Audio Teleportation* couples the location of the user's hearing with this target so they can auditorily preview the distant location. The user's audio perception is reset upon execution of the teleportation. Our technique presents a floating head at the target location that is vertically and rotationally aligned with the user's head to facilitate the anchoring of their auditory viewpoint. Our technique prioritizes translating the auditory viewpoint in the horizontal plane since horizontal movement often provides more useful information in environment-wide search tasks, particularly for locating target items (e.g., [58]). We maintain the rotational alignment of the auditory viewpoint since prior work suggests that rotationally-incongruent reference frames impair spatial cognition [37].

3.2 *Audio Cone*

Audio Cone was designed to amplify sounds from a particular direction, analogous to a flashlight or binoculars in the visual domain. It also draws inspiration from prior distant object selection techniques like the quad cone [39, 44] to enable the perception of otherwise inaudible distant sounds.

To use *Audio Cone*, users project a cone-shaped beam that increases the volume of the sound objects it encompasses. The beam originates from the user's hand, resembling a flashlight to provide a familiar experience and simple control. Users can leverage this information to determine the approximate direction to navigate, such as towards search targets.

3.3 *Ninja Ears*

Ninja Ears draws inspiration from prior work on *Ninja Cursors* [38] and *Ninja Hands* [57]. It generates a grid of distributed floating heads, representing potential audio perspectives from which the user can listen. However, in contrast to *Ninja Cursors* and *Ninja Hands*, which are designed to support simultaneous control of cursors and hands in multiple locations, *Ninja Ears* only allows activating one audio perspective at a time. This design decision was motivated by prior research that suggests spatial cues do not play a strong role in segmenting sounds when there are multiple sounds coming from multiple locations [46], which suggests that multiple active *Ninja Ears* will likely be more confusing than helpful.

Users use a bubble selector to choose which auditory viewpoint to listen from. Similar to *Audio Teleportation*, the orientations of the distant audio perspectives in *Ninja Ears* are always aligned with the user's head to avoid adverse impacts on spatial cognition from rotational incongruencies [37]. Likewise, the audio perception is reset to align with the user's normal viewpoint upon teleportation.

Ninja Ears is similar to *Audio Teleportation* as it allows users to hear from a different location than their visual viewpoint. However, unlike *Audio Teleportation*, *Ninja Ears* decouples auditory perspective selection from teleportation, letting users choose from several predetermined hearing locations. While this may increase the user's freedom in auditory exploration, it sacrifices the ability to select the exact location of their hearing.

3.4 *Boom Mic*

Boom Mic mimics a physical approach to capturing sounds at a distance. The technique is effectively a virtual instantiation of its real counterpart, which is a microphone mounted on the end of a boom pole. We represent the *Boom Mic* virtually as a head fixed to a length-adjustable stick, to which the user's auditory perspective is attached when activated.

3.5 Implementation

We implemented prototypes for *Audio Teleportation*, *Audio Cone*, *Ninja Ears*, and *Boom Mic* in Unity 2022.3.19f1 for the Quest 2 [35]. HRTF spatial audio was synthesized using the Meta XR Audio SDK version 62.0.0. All techniques were operated with a Quest 2 controller. As an environment to host the techniques, we designed a 30 $m \times 21$ m virtual office (Figure 3). The environment was populated with typical office objects, such as desks and chairs, all with low-resolution textures and low-polygon meshes.

The prototype application supports locomotion through *Teleportation*. We used a standard *Teleportation* implementation, which includes raycast-based location selection and joystick-based snap turns. All *Teleportation* controls were managed with the right controller. *Audio Teleportation* directly integrated with the *Teleportation* controls, requiring no additional input. The additional controls needed for *Audio Cone*, *Ninja Ears*, and *Boom Mic* were all attached to the user's left controller.

The direction of the *Audio Cone* was mapped to the direction of the controller. Users could additionally adjust its radius continuously within a range of $(0, 180°)$ with the controller joystick.

We set *Ninja Ears* to spawn a grid of 9×6 of "ears" within the prototype environment. Users can initially activate an audio perspective by pressing down on their controller joystick. This action automatically selects the closest floating head, indicated with a bubble selector. They can then toggle which head is selected by pushing the joystick in a desired direction based on their current perspective.

Lastly, the *Boom Mic* is attached to their left controller, aligned with the direction of the boom stick, and positioned as though it is held in their left hand. Users can adjust the length of its extension with the joystick. Inspired by the Go-Go interaction technique [53], the length of the boom pole is calculated using a quadratic function for efficient reaching across the space.

4 Experiment

In order to understand the benefits and limitations of our audio interaction techniques for navigation, we designed a search task that required participants to find all gems present in a virtual office environment. This task was adapted from a previous experiment reported in the literature [36, 40].

Our study involved 25 participants. We compare the effect of five INTERACTION TECHNIQUES on participants' search performance (e.g., *search time*), behavior (e.g., *distance*), and subjective preferences (e.g., *usability*).

4.1 Design

The experiment followed a within-subject design with one independent variable with five levels: INTERACTION TECHNIQUE (*Teleportation*, *Audio Teleportation*, *Audio Cone*, *Ninja Ears*, *Boom Mic*). In the *Audio Teleportation*, *Audio Cone*, *Ninja Ears*, and *Boom Mic* conditions, participants could use the respective audio interaction techniques in addition to locomotion through *Teleportation*.

To mitigate ordering effects, we counterbalanced the INTERACTION TECHNIQUE order using a Latin Square, resulting in 10 possible orders. For the first 20 participants, we used all 10 possibilities twice. For the last five participants, we randomly selected five out of the existing possibilities. We analyzed ORDER effects (5 levels, i. e., condition orders, between-subject) and did not find a significant effect of order on any of the dependent variables (all $p > .05$).

4.2 Task

Participants performed a gem search task. They needed to find 12 gems in the space, using the INTERACTION TECHNIQUE provided. The task represents a common operation in VR, where the user needs to locate and collect objects at various distant locations [58].

For the search task, one gem was generated at a time. Gems were accompanied by a persistent sound effect, designed to mimic the auditory cues commonly used for collectibles or points of interest in video games. We adjusted the volume of the accompanying sound to only be audible within a range of 7.5 *m* (i. e., one-quarter of the virtual office's length), once again inspired by the proximity-based activation of such cues in gaming environments.

For each condition, gems were randomly placed in 12 of 36 possible locations (Figure 3). We positioned half of these gems in

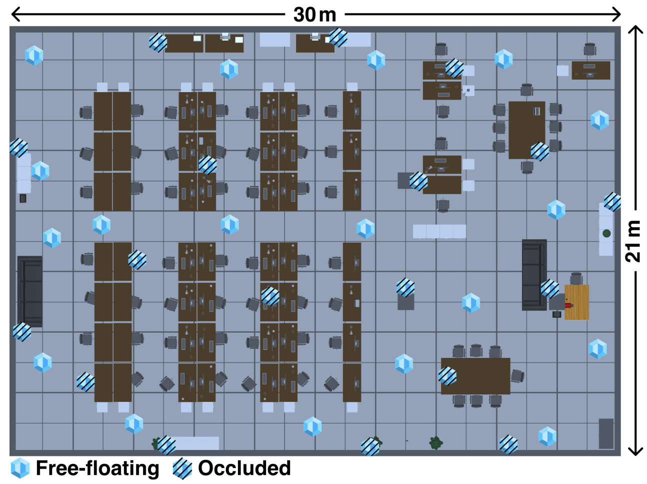

Figure 3: Virtual environment of our experiment. Participants performed a gem search task in an office scene. Each condition consisted of 6 free-floating and 6 occluded gems, randomly sampled from the indicated locations.

open spaces, away from any nearby objects within the scene, and the other half in locations within 0.5 *m* of an object, partially obscured from at least one viewing angle (e.g., beneath a table, behind a plant, within a trash can). Participants pick up gems by touching them with colliders around their controllers (10 *cm* diameter). One trial consists of collecting a single gem. Audio feedback is provided for successful collection, after which a new gem is generated in a different location. We set the minimum distance between consecutive gem locations to be 2 *m*, to require some exploration of the environment in each trial.

4.3 Participants

We recruited 25 participants via snowball sampling starting from university message groups and social networks. We chose this sample size based on a similar experiment conducted by Kumaran et al. [40]. Participants had to be 18-70 years old, with no known visual or hearing impairments that would compromise their experience of a VR application with spatial audio. Participants were compensated $25 for their time.

In the pre-questionnaire, we asked participants to report their demographic information (age, gender), prior experience with VR (7-point Likert scale, from 1-none to 7-expert), frequency of playing video games (1-never, 2-once every 2 or 3 months, 3-once every month, 4-once every 2 weeks, 5-at least once a week), and their level of alertness using the Stanford Sleepiness Scale [63] (from 1-active, vital, alert, or wide awake to 7-sleep onset soon). The median responses from participants (age: $M = 26$ *years old*, $SD = 4$ *years old*; 15 female, 10 male) are: VR experience = 3, gaming frequency = 2, and level of alertness = 2. No participants reported any visual or hearing impairments, and we observed no performance variations in the results to suggest otherwise.

4.4 Apparatus

Participants performed all tasks in a 2 *m* diameter circular area inside a quiet, designated experimental space (Figure 4). They were equipped with a Quest 2 headset and AKG Pro K121 Studio over-ear headphones. The experiment ran on an Intel Core i7-12700H CPU 2.30 GHz computer with 16 GB of RAM, supported by an NVIDIA GeForce RTX 3060 GPU. We conducted our experiment using a Quest 2, as it is one of the most popular VR headsets on the market [35], and the AKG Pro headphones, as it is a professional standard in studios [66]. The Meta XR Audio SDK uses a generic

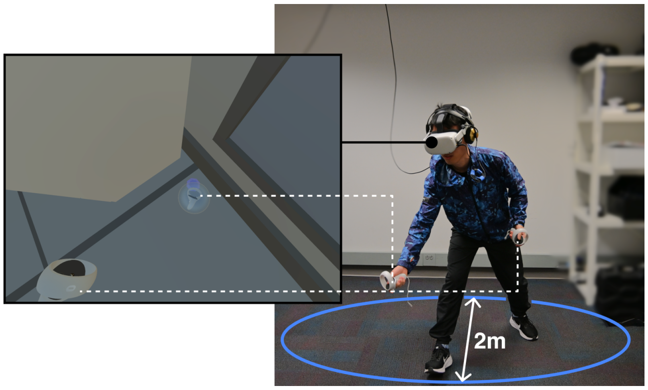

Figure 4: Apparatus of our experiment. Participants were equipped with a Meta Quest 2 headset and AKG Pro K121 Studio over-ear headphones to perform a search task in VR.

HRTF for rendering spatial audio, which Berger et al. [5] suggest is potentially sufficient for VR applications. We discuss the potential value of using a personalized HRTF instead in Section 6.4.

We leverage the same 30 $m \times 21$ m virtual scene described in Section 3.5 to host the experiment. We determined the dimensions of the scene based on prior navigation experiments (e.g., [40, 58]).

4.5 Procedure

After obtaining informed consent from the participants, the experimenter first introduced them to the study, equipment, and recorded data. Participants then completed a pre-questionnaire to assess their experience with VR and video games, their state of alertness, and their demographic profile.

Subsequently, participants proceeded through the conditions of our within-subject design. Before performing the evaluated search task in each condition, participants completed four trials (with two open and two occluded gems) as training to become accustomed to the interaction technique controls and the virtual environment. Participants then proceeded onto the recorded trials. After each condition, participants reported on several subjective metrics (e.g., *usability*) in a post-condition survey. Between conditions, participants were allowed to rest for as long as they preferred. The study concluded with an exit survey asking participants to rank the interaction techniques by preference, followed by a semi-structured interview that inquired about their preferences, suggestions for improvement, and envisioned applications of the techniques. All survey items and interview questions are provided in the supplementary materials. The entire procedure took approximately 75 *min* per participant.

4.6 Measures

We analyzed the following metrics:

Search task performance and behavior: We recorded the *search time*, *travel distance*, and *number of teleportations* for each trial. To account for differences in distance between trials, we also computed *normalized search time*, *distance*, and *teleportation count* metrics. This normalization involved dividing each value by the distance between the starting and target gem positions.

Self-reported Metrics: After each condition, we asked participants to evaluate the *usability* of the INTERACTION TECHNIQUE using the short version of the User Experience Questionnaire (UEQ-S) [59]. Since several of the audio interaction techniques involved displacing participants' auditory perspective from their visual perspective, we were also interested in how they influenced participants' sense of *presence* and *embodiment*. We used a subset of the questions in the Igroup Presence Questionnaire (IPQ) [60] to assess presence, specifically focusing on those associated with a general sense of *presence* (G1) and *spatial presence* (SP1-SP5). We used a subset of the questions in Gonzalez-Franco and Peck's Avatar Embodiment questionnaire [28] to assess *embodiment*, specifically focusing on those associated with *ownership* (Q1-Q3), *agency* (Q6-Q9), and *location* (Q14-Q15). Lastly, we asked participants to rate their level of motion sickness after each condition (1-none, 7-severe).

In addition to the post-condition questionnaire questions, we asked participants to rank the INTERACTION TECHNIQUE by *preference* at the end of the experiment.

5 RESULTS

We analyzed the interval data using a one-way repeated measures ANOVA. When the equal variances assumption was violated (Mauchly's test $p < .05$), we corrected the degrees of freedom using Greenhouse-Geisser. When the assumption about the normality of residuals and homogeneity was violated (Shapiro-Wilk test $p < .05$), we either transformed the data using a log function or analyzed them using the Aligned Rank Transform (ART) [71]. Ordinal data (questionnaire ratings) was analyzed using ART. For each data value, the *participant* was considered a random factor and the INTERACTION TECHNIQUE was set as the independent variable. Pairwise comparisons using the Bonferroni correction were used to follow up significant main effects. The statistical analysis was performed using IBM SPSS 29 [32].

5.1 Search Task Performance & Behavior

All task performance and behavior results are shown in Figure 5.

Time: We transformed *time* using a log function for the analysis. The ANOVA analysis showed a significant main effect of INTERACTION TECHNIQUE ($F_{1,4} = 14.58$, $p < .001$, $\eta_p^2 = .38$). Participants were slower with *Boom Mic* than *Audio Cone* (by 68%, $p < .001$), *Audio Teleportation* (by 64%, $p < .001$), *Ninja Ears* (by 35%, $p < .01$), and *Teleportation* (by 30%, $p < .001$). They were 20% faster with *Audio Cone* than *Ninja Ears* ($p = .03$).

Distance: The ART analysis showed a main effect of INTERACTION TECHNIQUE ($F_{1,4} = 17.13$, $p < .001$, $\eta_p^2 = .42$). Participants moved more with *Teleportation* than *Audio Cone* (by 109%, $p < .001$), *Audio Teleportation* (by 91%, $p < .001$), *Ninja Ears* (by 67%, $p < .001$), and *Boom Mic* (by 99%, $p < .01$).

Number of teleportations: The ART analysis showed a main effect of INTERACTION TECHNIQUE ($F_{1,4} = 26.55$, $p < .001$, $\eta_p^2 = .53$). Participants teleported more with *Teleportation* (all $p < .001$) than *Audio Cone* (by 193%), *Audio Teleportation* (by 181%), *Ninja Ears* (by 209%), and *Boom Mic* (by 163%).

Normalized time: We transformed *normalized time* using a log function for the analysis. The ANOVA analysis showed a significant main effect of INTERACTION TECHNIQUE ($F_{1,4} = 8.07$, $p < .001$, $\eta_p^2 = .25$). Accounting for distance differences between trials, participants were faster with *Audio Cone* than *Boom Mic* ($p < .001$) and *Ninja Ears* ($p = .03$) by 37% and 22%, respectively. They were also 31% faster with *Audio Teleportation* than *Boom Mic* ($p < .001$).

Normalized distance: The ART analysis showed a main effect of INTERACTION TECHNIQUE ($F_{1,4} = 17.33$, $p < .001$, $\eta_p^2 = .42$). Accounting for distance differences between trials, participants moved more with *Teleportation* (all $p < .001$) than *Audio Cone* (by 124%), *Audio Teleportation* (by 96%), *Ninja Ears* (by 112%), and *Boom Mic* (by 79%).

Normalized number of teleportations: The ART analysis showed a main effect of INTERACTION TECHNIQUE ($F_{1,4} = 22.21$, $p < .001$, $\eta_p^2 = .48$). Accounting for distance differences between trials, participants teleported more with *Teleportation* (all $p < .001$) than *Audio Cone* (by 195%), *Audio Teleportation* (by 152%), *Ninja Ears* (by 191%), and *Boom Mic* (by 153%).

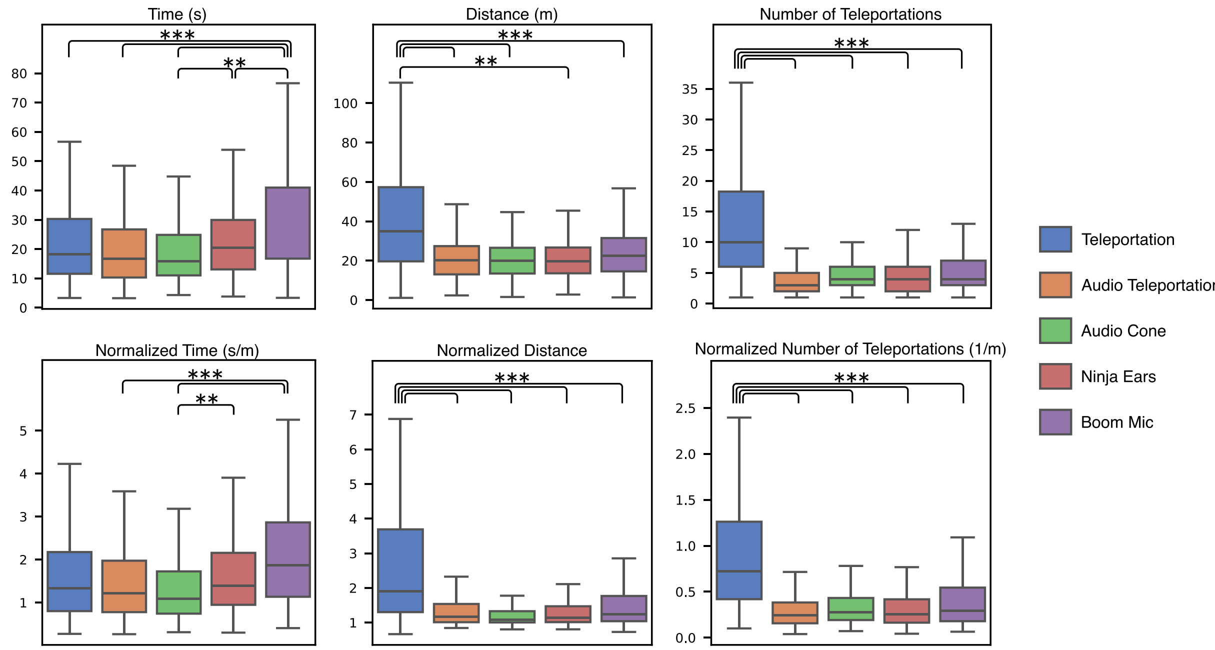

Figure 5: Effect of INTERACTION TECHNIQUE on *time*, *distance*, *teleportations*, *normalized time*, *normalized distance*, and *normalized teleportations*. Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$.

5.2 Subjective Ratings

The following analyses were performed using ART. Significant main effects were found in the measures associated with *usability* and *preference*, whereas no significant effects were observed in measures related to *presence*, *embodiment*, and *motion sickness*.

Usability: The ART analysis revealed a main effect of INTERACTION TECHNIQUE on the cumulative metrics of overall *usability* ($F_{1,4} = 13.22$, $p < .001$, $\eta_p^2 = .36$), *pragmatic quality* ($F_{1,2.85} = 12.25$, $p < .001$, $\eta_p^2 = .34$), and *hedonic quality* ($F_{1,4} = 10.28$, $p < .001$, $\eta_p^2 = .3$), as well as on each individual dimension of the UEQ-S (all $p < .001$).

Participants considered *Audio Cone* ($mdn(median) = 2.25$) to be more usable than *Teleportation* ($mdn = 0.38$, $p < .001$), *Ninja Ears* ($mdn = 1.83$, $p = .04$), and *Boom Mic* ($mdn = 1.5$, $p < .001$). They also considered *Audio Teleportation* ($mdn = 2.25$) to be more usable than *Teleportation* ($p < .001$) and *Boom Mic* ($p < .01$). They regarded *Audio Cone* ($mdn(median) = 2.5$) as more *pragmatic* than *Teleportation* ($mdn = 0.75$, $p = .02$), *Ninja Ears* ($mdn = 1.75$, $p < .001$), and *Boom Mic* ($mdn = 1$, $p < .001$). They also regarded *Audio Teleportation* ($mdn = 2.75$) as more *pragmatic* than *Teleportation* ($p < .01$), *Ninja Ears* ($p = .02$), and *Boom Mic* ($p < .001$). Last, participants attributed a higher *hedonic quality* to *Audio Cone* ($mdn = 2.25$, $p < .01$), *Audio Teleportation* ($mdn = 2.25$, $p < .001$), *Ninja Ears* ($mdn = 2$, $p < .01$), *Boom Mic* ($mdn = 1.75$, $p = .05$) than *Teleportation* ($mdn = -0.25$). We show significant post-hoc comparisons for each UEQ-S dimension in Figure 6.

Technique preference: The ART analysis revealed a main effect of INTERACTION TECHNIQUE ($F_{1,3.06} = 18.57$, $p < .001$, $\eta_p^2 = .43$). Participants preferred *Audio Cone* ($mdn = 1$, all $p < .001$) over *Teleportation* ($mdn = 5$), *Ninja Ears* ($mdn = 3$), and *Boom Mic* ($mdn = 4$). They also preferred *Audio Teleportation* ($mdn = 2$) over *Teleportation* ($p < .001$), *Ninja Ears* ($p = .02$), and *Boom Mic* ($p < .001$).

5.3 Concluding Interview

Exit interviews with participants were transcribed and analyzed through affinity diagramming, revealing four themes.

Supporting search with audio. Since the *Teleportation* technique provided "limited audio cues," participants associated their experience of using it with a "sense of uncertainty" ($N = 15$). The proposed interaction techniques addressed this limitation in two ways.

First, participants primarily attributed the value of *Audio Cone* to its support for identifying the general *direction* of a target relative to themselves ($N = 10$). By pointing the user in approximately the right direction, users can effectively "walk along a line" until they reach the target object (P7). *Audio Cone*, however, provides limited information about the exact *location* of the target ($N = 4$), which is, in contrast, what participants appreciated about *Ninja Ears* and *Teleportation* ($N = 8$). By allowing users to quickly toggle between different auditory perspectives, they can efficiently identify locations that auditorily signal the presence of the target.

Participants also reported that while both approaches offer advantages in locating distant targets, for closer targets, they sometimes adversely impacted search performance by creating "local confusion" ($N = 5$), wherein participants overlooked nearby targets (e.g., P3: "I was extending the *Ninja Ears* everywhere, but the gem was just right next to me").

Integration. One aspect of the interaction techniques that significantly affected participants' user experience was the extent to which they were integrated with the locomotion controls. For instance, most participants ($N = 15$) appreciated *Audio Teleportation* for how it seamlessly "merges audio cues in teleportation" (P0). In contrast, participants reported that techniques like *Ninja Ears* and *Boom Mic* separate the processes of auditory surveying and locomotion, which in turn makes control more "tedious" ($N = 12$).

Visual feedback. Participants also reported that the visual cues provided as part of the interaction techniques affected their task performance. According to P14 and P24, the floating head indicator used in *Audio Teleportation* helped them make sense of the position and orientation of their auditory perspective. In contrast, several participants ($N = 4$) remarked that the bubble selector used in *Ninja Ears* made the position and orientation of their auditory perspective more difficult to identify, making it more challenging to make sense of the audio cues they received.

Applications. Last, participants cited a wide range of potential applications for our interaction techniques. Besides object search in

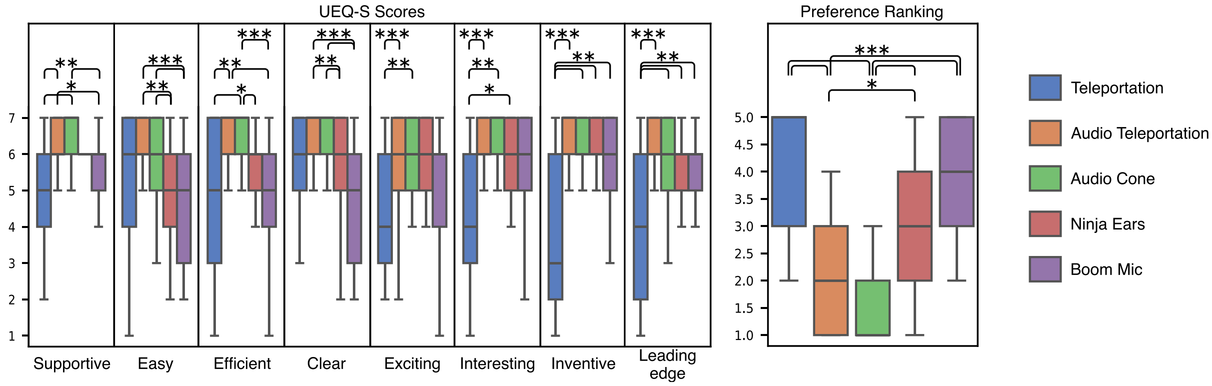

Figure 6: Effect of INTERACTION TECHNIQUE on usability and preference ratings. Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$.

virtual environments (N=12), participants also envisioned its usage in social VR settings (N=7), gameplay (N=12), accessibility support (N=5), and passive auditory environment experiences (N=7).

6 Discussion

Our results show that supplementing search with interaction techniques for augmenting *auditory perception* can be beneficial. The techniques significantly reduced participants' teleportation count and distance, which may help reduce disorientation and motion sickness over time. Participants preferred *Audio Cone* and *Audio Teleportation* over the baseline teleportation technique, attributing to both a higher pragmatic and hedonic quality. However, not all proposed interaction techniques were equally effective or appreciated. *Boom Mic* increased search time compared to the baseline, was preferred the least, and was perceived most negatively in terms of usability. Last, we found comparable levels of presence, embodiment, and motion sickness across all techniques.

6.1 Search Task Performance & Behavior

Besides *Boom Mic*, the proposed techniques generally yielded similar search times compared to the baseline. However, all techniques reduced the teleportation count and distance participants traveled. This reduction in teleportation and travel distance can potentially reduce motion sickness [13], improve spatial understanding [55], and enhance the continuity and immersion of the VR experience [64]. Moreover, it may indicate more efficient navigation.

Instead of arbitrarily teleporting around, participants could leverage the auditory interaction techniques to move more deliberately (Figure 7). Despite these benefits, the comparable search times suggest that visual cues still dominate the search task given our experimental settings, aligning with prior research [16].

Boom Mic resulted in the longest search times. This is mostly because it caused confusion when searching locally and presented challenges in aligning target locations with the *Boom Mic* (see Section 5.3). Participants were also faster with *Audio Cone* than *Ninja Ears*. *Audio Cone* has a directional range that enables participants to search through the environment with more confidence. In contrast, *Ninja Ears* allows selection between different point ranges, which ended up complicating the existing search task flow. To further understand these aspects, future work may consider conducting systematic investigations into how different ranges of audio interaction techniques would impact search efficiency and task loads, as well as how they can be adapted for different sound conditions.

6.2 Self-reports

Participants ranked *Audio Cone* and *Audio Teleportation* highest in terms of preference, regarding them as the most usable, with the highest pragmatic and hedonic quality. Participants' responses for

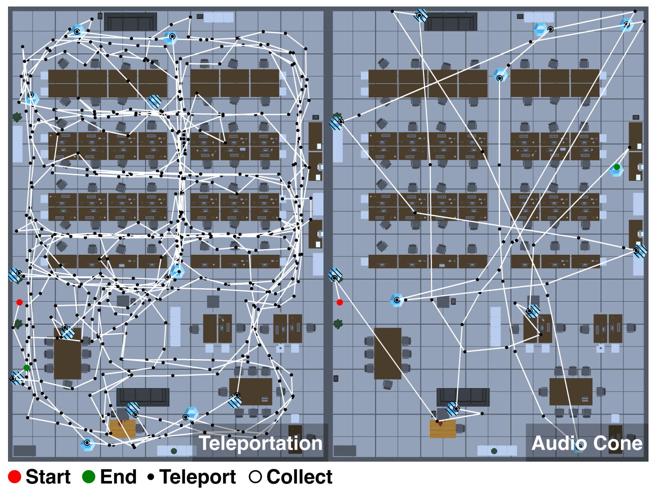

Figure 7: An example comparison of navigation paths between *Teleportation* and *Audio Cone* (P8).

the individual dimensions of the UEQ-S suggest the techniques are perceived as supportive, easy to use, efficient, and clear. The techniques were also regarded as exciting, interesting, and inventive. As discussed in Section 5.3, *Audio Cone* was liked mostly for its intuitiveness, flexibility, and easiness of operation. While it did not assist users in determining the precise location and depth of targets, its ability to provide direction was considered helpful. *Audio Teleportation* was appreciated mostly for its seamless integration of audio interaction and teleportation controls, but its limited range was found problematic for larger environments.

Preference. *Boom Mic* and *Ninja Ears* were ranked comparatively lower in terms of preference. One of the core limitations of the two techniques is that they are less pragmatic, though they offer a similar hedonic quality as *Audio Cone* and *Audio Teleportation*. As revealed in the concluding interview, participants found both techniques to be "cool" and "playful", but less usable. Participants indicated that they might be better suited for other usage scenarios (e.g., *Boom Mic* for tasks involving vertical resolutions, *Ninja Ears* for sound monitoring tasks).

Presence & embodiment. Interestingly, we found comparable levels of presence and embodiment across our techniques. The techniques effectively separate the users' visual and auditory perspectives, in contrast to how they are typically collocated in physical reality. Being in multiple locations at once might have affected the user's sense of presence and embodiment, but these differences were not reflected in the results. For the sense of presence in the virtual environment, one conjecture could be that although audio interactions altered the way participants perceive sound, their impact was minor - with only one audio source, i. e., the gem, in the environment - whereas the overall experience was predominantly visual-based. Future work may consider designing more audio-heavy environment setups as well as further investigating the believability of beyond-real audio interaction techniques.

Motion sickness. Last, we found comparable levels of motion sickness across our techniques, notably in spite of *Teleportation* resulting in more teleportations and locomotion distance. However, the average rating participants provided across different conditions was 2.07 ($SD = 1.36$), close to the lower bound, which suggests that we may be observing floor effects. This aligns with prior work that suggests *Teleportation* induces low motion sickness overall [54]. To better understand how different auditory interaction techniques might impact motion sickness, future studies could consider adopting more complex environment designs.

6.3 Implications for Design

Based on our findings, we have gained the following insights into designing audio interaction techniques in VR:

- Audio interaction techniques (e.g., *Audio Teleportation*, *Audio Cone*) offer an alternative approach to exploring virtual environments that may aid decisions relating to navigation and reduce visual movements and viewpoint changes, thereby enhancing the user experience.
- Audio interaction techniques should balance providing *directional* and *distance* information based on the context.
- Techniques that integrate seamlessly with existing locomotion controls are generally preferred (e.g., *Audio Teleportation*). The trade-off between added functionality and control overhead must be carefully considered.
- Visual feedback is important for helping users make sense of their altered auditory perspective, thereby improving the usability of audio interaction techniques.

6.4 Applications and Future Work

Our study investigated four audio interaction techniques and a baseline teleportation technique for a search task. We believe our techniques have a variety of potential applications, as supported by our study participants who cited scenarios ranging from social VR to accessibility support. For instance, in social VR settings, we envision Audio Cone enabling users to selectively focus on their conversation partner and Ninja Ears offering the ability to converse at a distance. In gaming contexts, all techniques may serve as interesting mechanisms to augment gameplay (e.g., power-ups in a puzzle game or first-person shooter). Furthermore, while not the focus of our present study, exploring how auditory interaction techniques can promote accessibility in VR applications may be another valuable research direction. For instance, future work may consider extending the current techniques into a plug-and-play toolkit to accommodate different hearing abilities.

However, for these aforementioned applications to come to fruition, further research is required. For starters, we acknowledge that although our exploratory study yielded some interesting insights, it represents more of a starting point than a comprehensive investigation. Our work helped highlight a few key dimensions of the design space for auditory interaction techniques in search, but a more systematic investigation will be valuable for understanding its various parameters and optimizing their design. Additionally, while our current study involved searching for consecutive targets in soundless environments, studying the efficacy of our techniques in less controlled contexts, including multiple targets and sound sources, would broaden the applicability of our results.

Furthermore, in our experiment, we focused on evaluating search performance; however, in navigating virtual environments, cognitive mapping and maintaining spatial orientation are equally important dimensions. Future work may therefore consider investigating how augmenting one's auditory experience of an environment, particularly through separating one's audio and visual perspectives, influences spatial cognition and navigation strategies. More broadly, a controlled experiment on the effect of the user's auditory perspective on their sense of embodiment may make for an interesting direction of future research.

Finally, our experiment currently relies on virtually synthesized spatial audio based on a generic HRTF. While this is representative of current VR experiences and suggested by prior research to be a reasonable design decision [5], future studies may consider exploring the value of using a personalized HRTF instead, particularly in the context of audio interaction techniques that augment users' auditory experiences of the virtual environment, rather than just experiences that replicate reality.

7 Conclusion

In this paper, we explored auditory interaction techniques for navigating virtual environments. We developed four approaches to augment users' perception: *Audio Teleportation*, *Audio Cone*, *Ninja Ears*, and *Boom Mic*. Our comparative user study with 25 participants revealed that these audio techniques can potentially lead to more efficient search behaviors in VR, as evident by reductions in teleportation usage and movement distances, while also enhancing user satisfaction. Our results also showed that the specific design of the audio interaction technique affected its effectiveness and user experience, with dimensions like the familiarity of its design, the extent of its integration with the locomotion controls, and the visual feedback all playing a significant role. Overall, our study provides insights into the effective design of future audio interaction techniques for search and navigation in VR applications.

Acknowledgments

We thank all involved peers, participants, and anonymous reviewers, especially Radha Kumaran for her input on the task design. This work was supported in part by the Croucher Foundation.

References

[1] P. Abtahi, S. Q. Hough, J. A. Landay, and S. Follmer. Beyond being real: A sensorimotor control perspective on interactions in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517706 2

[2] M. Al Zayer, P. MacNeilage, and E. Folmer. Virtual locomotion: a survey. *IEEE transactions on visualization and computer graphics*, 26(6):2315–2334, 2018. 1

[3] P. Bala, R. Masu, V. Nisi, and N. Nunes. " when the elephant trumps" a comparative study on spatial audio for orientation in 360º videos. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019. 1

[4] B. B. Bederson. Audio augmented reality: a prototype automated tour guide. In *Conference companion on Human factors in computing systems*, pp. 210–211, 1995. 2

[5] C. C. Berger, M. Gonzalez-Franco, A. Tajadura-Jiménez, D. Florencio, and Z. Zhang. Generic hrtfs may be good enough in virtual reality. improving source localization through cross-modal plasticity. *Frontiers in neuroscience*, 12:21, 2018. 5, 8

[6] L. Berger and K. Wolf. Wim: fast locomotion in virtual reality with spatial orientation gain & without motion sickness. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pp. 19–24, 2018. 2

[7] M. Billinghurst, J. Bowskill, M. Jessop, and J. Morphett. A wearable spatial conferencing space. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No. 98EX215)*, pp. 76–83. IEEE, 1998. 2

[8] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg. Earcons and icons: Their structure and common design principles. *Human–Computer Interaction*, 4(1):11–44, 1989. 2

[9] D. A. Bowman, D. Koller, and L. F. Hodges. Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*, pp. 45–52. IEEE, 1997. 2

[10] D. A. Bowman, D. Koller, and L. F. Hodges. A methodology for the evaluation of travel techniques for immersive virtual environments. *Virtual reality*, 3(2):120–131, 1998. 2

[11] E. Bozgeyikli, A. Raij, S. Katkoori, and R. Dubey. Point & teleport locomotion technique for virtual reality. In *Proceedings of the 2016 annual symposium on computer-human interaction in play*, pp. 205–216, 2016. 1, 2, 3

[12] S. Burigat and L. Chittaro. Navigation in 3d virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human-Computer Studies*, 65(11):945–958, 2007. 1

[13] E. Chang, H. T. Kim, and B. Yoo. Virtual reality sickness: a review of causes and measurements. *International Journal of Human–Computer Interaction*, 36(17):1658–1682, 2020. 7

[14] R.-C. Chang, C.-S. Hung, B.-Y. Chen, D. Jain, and A. Guo. Sound unblending: Exploring sound manipulations for accessible mixed-reality awareness. *arXiv preprint arXiv:2401.11095*, 2024. 2

[15] R.-C. Chang, C.-H. Ting, C.-S. Hung, W.-C. Lee, L.-J. Chen, Y.-T. Chao, B.-Y. Chen, and A. Guo. Omniscribe: Authoring immersive audio descriptions for 360 videos. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–14, 2022. 2

[16] T. Chen, Y.-S. Wu, and K. Zhu. Investigating different modalities of directional cues for multi-task visual-searching scenario in virtual reality. In *Proceedings of the 24th ACM symposium on virtual reality software and technology*, pp. 1–5, 2018. 7

[17] G. D. Clemenson, A. Maselli, A. J. Fiannaca, A. Miller, and M. Gonzalez-Franco. Rethinking gps navigation: creating cognitive maps through auditory clues. *Scientific reports*, 11(1):7764, 2021. 2

[18] D. Dewez, L. Hoyet, A. Lécuyer, and F. Argelaguet. Studying the inter-relation between locomotion techniques and embodiment in virtual reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 452–461. IEEE, 2020. 2

[19] J. Dodiya and V. N. Alexandrov. Use of auditory cues for wayfinding assistance in virtual environment: music aids route decision. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pp. 171–174, 2008. 2

[20] A. O. S. Feiner. The flexible pointer: An interaction technique for selection in augmented and virtual reality. In *Proc. UIST*, vol. 3, pp. 81–82, 2003. 2

[21] L. Fialho, J. Oliveira, A. Filipe, and F. Luz. Soundspace vr: spatial navigation using sound in virtual reality. *Virtual Reality*, 27(1):397–405, 2023. 2

[22] K. Franinovic and S. Serafin. *Sonic interaction design*. Mit Press, 2013. 2

[23] Z. Gao, H. Wang, G. Feng, and H. Lv. Exploring sonification mapping strategies for spatial auditory guidance in immersive virtual environments. *ACM Transactions on Applied Perceptions (TAP)*, 19(3):1–21, 2022. 2

[24] W. W. Gaver. Auditory icons: Using sound in computer interfaces. *ACM SIGCHI Bulletin*, 19(1):74, 1987. 2

[25] W. W. Gaver. Auditory interfaces. In *Handbook of human-computer interaction*, pp. 1003–1041. Elsevier, 1997. 2

[26] M. Geronazzo, L. S. Vieira, N. C. Nilsson, J. Udesen, and S. Serafin. Superhuman hearing - virtual prototyping of artificial hearing: a case study on interactions and acoustic beamforming. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1912–1922, 2020. doi: 10.1109/TVCG.2020.2973059 2, 3

[27] R. G. Golledge. *Wayfinding behavior: Cognitive mapping and other spatial processes*. JHU press, 1999. 2

[28] M. Gonzalez-Franco and T. C. Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018. 5, 1

[29] T. Grossman and R. Balakrishnan. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 281–290, 2005. 2

[30] G. Haas, E. Stemasov, M. Rietzler, and E. Rukzio. Interactive auditory mediated reality: Towards user-defined personal soundscapes. In *Conference on Designing Interactive Systems*, pp. 2035–2050, 2020. 2

[31] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. A survey of design issues in spatial input. In *Proceedings of the 7th annual ACM symposium on User interface software and technology*, pp. 213–222, 1994. 2

[32] IBM Corp. *IBM SPSS Statistics for Windows, Version 29.0*. IBM Corp., Armonk, NY, 2021. 5

[33] D. Jain, S. Junuzovic, E. Ofek, M. Sinclair, J. Porter, C. Yoon, S. Machanavajhala, and M. Ringel Morris. A taxonomy of sounds in virtual reality. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference*, pp. 160–170, 2021. 2

[34] S. Jul and G. W. Furnas. Navigation in electronic worlds: a chi 97 workshop. *ACM SIGCHI Bulletin*, 29(4):44–49, 1997. 2

[35] J. W. Kelly, T. A. Doty, M. Ambourn, and L. A. Cherep. Distance perception in the oculus quest and oculus quest 2. *Frontiers in Virtual Reality*, 3:850471, 2022. 3, 4

[36] Y.-J. Kim, R. Kumaran, E. Sayyad, A. Milner, T. Bullock, B. Giesbrecht, and T. Höllerer. Investigating search among physical and virtual objects under different lighting conditions. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3788–3798, 2022. doi: 10.1109/TVCG.2022.3203093 4

[37] R. L. Klatzky. Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge*, pp. 1–17. Springer, 1998. 3

[38] M. Kobayashi and T. Igarashi. Ninja cursors: using multiple cursors to assist target acquisition on large screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 949–958, 2008. 2, 3

[39] R. Kopper, F. Bacim, and D. A. Bowman. Rapid and accurate 3d selection by progressive refinement. In *2011 IEEE symposium on 3D user interfaces (3DUI)*, pp. 67–74. IEEE, 2011. 2, 3

[40] R. Kumaran, Y.-J. Kim, A. E. Milner, T. Bullock, B. Giesbrecht, and T. Höllerer. The impact of navigation aids on search performance and object recall in wide-area augmented reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581413 4, 5

[41] J. J. LaViola Jr. A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin*, 32(1):47–56, 2000. 2

[42] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017. 1, 2

[43] J. Lee, S. C. Ahn, and J.-I. Hwang. A walking-in-place method for virtual reality using position and orientation tracking. *Sensors*, 18(9):2832, 2018. 2

[44] J. Liang and M. Green. Jdcad: A highly interactive 3d modeling system. *Computers & graphics*, 18(4):499–506, 1994. 2, 3

[45] J. M. Loomis, R. G. Golledge, and R. L. Klatzky. Navigation system for the blind: Auditory display modes and guidance. *Presence*, 7(2):193–203, 1998. 1, 2

[46] A. Lotto and L. Holt. Psychology of auditory perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5):479–489, 2011. 3

[47] D. Massiceti, S. L. Hicks, and J. J. van Rheede. Stereosonic vision: Exploring visual-to-auditory sensory substitution mappings in an immersive virtual reality navigation paradigm. *PloS one*, 13(7):e0199389, 2018. 2

[48] D. R. Montello. *Navigation.* Cambridge University Press, 2005. 2

[49] B. C. Moore. *An introduction to the psychology of hearing*. Brill, 2012. 2

[50] E. D. Mynatt, M. Back, R. Want, and R. Frederick. Audio aura: Lightweight audio augmented reality. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, pp. 211–

212, 1997. 2
[51] U. Neisser. Decision-time without reaction-time: Experiments in visual scanning. *The American journal of psychology*, 76(3):376–385, 1963. 1
[52] P. Panda, M. J. Nicholas, D. Nguyen, E. Ofek, M. Pahud, S. Rintel, M. Gonzalez-Franco, K. Hinckley, and J. Lanier. Beyond audio: Towards a design space of headphones as a site for interaction and sensing. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, pp. 904–916, 2023. 2
[53] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pp. 79–80, 1996. 2, 4
[54] A. Prithul, I. B. Adhanom, and E. Folmer. Teleportation in virtual reality; a mini-review. *Frontiers in Virtual Reality*, 2:730792, 2021. 2, 8
[55] R. A. Ruddle, E. Volkova, and H. H. Bülthoff. Walking improves your cognitive map in environments that are large-scale and large in extent. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(2):1–20, 2011. 7
[56] S. Russell, G. Dublon, and J. A. Paradiso. Hearthere: Networked sensory prosthetics through auditory augmented reality. In *Proceedings of the 7th Augmented Human International Conference 2016*, AH '16. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2875194.2875247 2
[57] J. Schjerlund, K. Hornbæk, and J. Bergström. Ninja hands: Using many hands to improve target selection in vr. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–14, 2021. 2, 3
[58] J. Schjerlund, K. Hornbæk, and J. Bergström. Ovrlap: Perceiving multiple locations simultaneously to improve interaction in vr. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3501873 2, 3, 4, 5
[59] M. Schrepp, A. Hinderks, and J. Thomaschewski. Design and evaluation of a short version of the user experience questionnaire (ueq-s). *International Journal of Interactive Multimedia and Artificial Intelligence, 4 (6), 103-108.*, 2017. 5, 1
[60] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001. 5, 1
[61] S. Serafin, F. Avanzini, A. De Goetzen, C. Erkut, M. Geronazzo, F. Grani, N. C. Nilsson, and R. Nordahl. Reflections from five years of sonic interactions in virtual environments workshops. *Journal of New Music Research*, 49(1):24–34, 2020. 2
[62] S. Serafin, M. Geronazzo, C. Erkut, N. C. Nilsson, and R. Nordahl. Sonic interactions in virtual reality: State of the art, current challenges, and future directions. *IEEE computer graphics and applications*, 38(2):31–43, 2018. 1, 2
[63] A. Shahid, K. Wilkinson, S. Marcu, and C. M. Shapiro. Stanford sleepiness scale (sss). *STOP, THAT and one hundred other sleep scales*, pp. 369–370, 2012. 4, 1
[64] M. Slater and M. Usoh. Body centred interaction in immersive virtual environments. *Artificial life and virtual reality*, 1(1994):125–148, 1994. 7
[65] J. L. Soler-Domínguez, C. de Juan, M. Contero, and M. Alcañiz. I walk, therefore i am: A multidimensional study on the influence of the locomotion method upon presence in virtual reality. *Journal of Computational Design and Engineering*, 7(5):577–590, 2020. 2
[66] G. Theile. Equalization of studio monitor headphones. In *Audio Engineering Society Conference: 2016 AES International Conference on Headphone Technology*. Audio Engineering Society, 2016. 4
[67] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks Jr. Walking¿ walking-in-place¿ flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 359–364, 1999. 2
[68] T. Van Gemert and J. Bergström. Evaluating vr sickness in vr locomotion techniques. In *2021 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*, pp. 380–382. IEEE, 2021. 2
[69] P. T. Wilson, W. Kalescky, A. MacLaughlin, and B. Williams. Vr locomotion: walking¿ walking in place¿ arm swinging. In *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry-Volume 1*, pp. 243–249, 2016. 1
[70] J. Wither and T. Hollerer. Evaluating techniques for interaction at a distance. In *Eighth International Symposium on Wearable Computers*, vol. 1, pp. 124–127. IEEE, 2004. 1
[71] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963 5
[72] Y. Yan, H. Liu, Y. Shi, J. Wang, R. Guo, Z. Li, X. Xu, C. Yu, Y. Wang, and Y. Shi. Conespeech: Exploring directional speech interaction for multi-person remote communication in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2647–2657, 2023. 2