An Empirical Evaluation of the Calibration of Auditory Distance Perception under Different Levels of Virtual Environment Visibilities

Wan-Yi Lin*
National Yang Ming Chiao Tung University

Rohith Venkatakrishnan*
University of Florida

Roshan Venkatakrishnan
University of Florida

Sabarish V. Babu
Clemson University

Christopher Pagano
Clemson University

Wen-Chieh Lin
National Yang Ming Chiao Tung University

2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR) | 979-8-3503-7402-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/VR58804.2024.00089

ABSTRACT

The perception of distance is a complex process that often involves sensory information beyond that of just vision. In this work, we investigated if depth perception based on auditory information can be calibrated, a process by which perceptual accuracy of depth judgments can be improved by providing feedback and then performing corrective actions. We further investigated if perceptual learning through carryover effects of calibration occurs in different levels of a virtual environment's visibility based on different levels of virtual lighting. Users performed an auditory depth judgment task over several trials in which they walked where they perceived an aural sound to be, yielding absolute estimates of perceived distance. This task was performed in three sequential phases: pretest, calibration, posttest. Feedback on the perceptual accuracy of distance estimates was only provided in the calibration phase, allowing to study the calibration of auditory depth perception. We employed a 2 (Visibility of virtual environment) $\times$ 3 (Phase) $\times$ 5 (Target Distance) multi-factorial design, manipulating the phase and target distance as within-subjects factors, and the visibility of the virtual environment as a between-subjects factor. Our results revealed that users generally tend to underestimate aurally perceived distances in VR similar to the distance compression effects that commonly occur in visual distance perception in VR. We found that auditory depth estimates, obtained using an absolute measure, can be calibrated to become more accurate through feedback and corrective action. In terms of environment visibility, we find that environments visible enough to reveal their extent may contain visual information that users attune to in scaling aurally perceived depth.

1 INTRODUCTION

Distance perception in virtual reality (VR) is a complex process influenced by various factors including visual, auditory, and haptic information [8]. A major challenge associated with VR experiences is the distorted perception of distance that occurs wherein objects are perceived to be closer than they actually are. This phenomenon is commonly referred to as distance compression, wherein users systematically underestimate distances in immersive virtual environments (IVEs) [27, 35, 43, 75]. This inaccuracy can potentially degrade the expected outcomes of experiences that require accurate depth perception seeing as how the perception of distance is closely linked to how users perceive aspects like size, scale, height, and speed. Industrial training applications, surgical simulations, and museum tour applications are examples of such experiences that may require higher levels of precision and perceptual accuracy [34, 47]. Tackling the depth compression issue is essential to address such concerns and for VR to be effective as an immersive technological solution to society's needs.

*both authors contributed equally to this research

In the real world, notwithstanding the dominance of vision as a sensory modality, people often leverage perceptual information from other sensory channels to make distance assessments [71]. With contemporary VR experiences often featuring high-quality spatial audio as a selling point for multisensory immersive experiences, it bodes well for researchers to study distance perception from the perspective of auditory information. Research on this front shows that visually perceived distance can be improved by altering auditory information presented along with the visual stimulus [25, 26, 35]. In terms of distance perception based solely on auditory information, researchers have found that users underestimate aurally perceived depth in medium and far-field ranges [42, 77]. Given the existence of similar systematic distortions in auditory depth perception, the research community continues to explore means to improve the accuracy of perceptual estimates of aurally specified distance.

Calibration/perceptual learning has been one of the most promising techniques to address inaccurate perceptions of depth [2, 7, 74]. Calibration occurs when users are provided feedback on their perceptual estimates along with the opportunity to make corrective actions toward making more accurate judgments [7, 56]. Studies have utilized calibration as a means to improve depth perception in VR, showing that reach estimates to visually presented targets significantly improve in accuracy after users undergo a calibration phase that provisions them with closed-loop feedback of their estimates made during a pretest phase [1, 18, 19]. Perceptual learning has also been studied with aurally perceived distances, with recent research showing the existence of calibrative carryover effects toward improving perceptual accuracy [39]. In this study, users performed a number of auditory distance estimation trials, lighting up from a series of light bulbs, in each trial, the bulb they felt most accurately represented how far they perceived a sound source to be. It was found that after undergoing a calibration phase with closed-loop feedback on their perceptual judgments, users' estimates were significantly more accurate in a posttest phase as compared to a pretest phase. It must be noted that this study's depth estimates involved a relative measure of distance using referent objects (lightbulbs) rather than an absolute egocentric measure of distance like reach estimates or blind walking. Given that absolute measures are generally more accurate than relative measures or verbal reports [45, 51], it bodes well for researchers to ascertain if calibration of aural distance perception is possible using an absolute measure.

Even in auditory depth perception - the perception of a target's distance based on aural information - visual information is of relevance. Real-world research on this front has shown that allowing users to see the room in which an auditory depth judgment task is going to be performed in an experiment produces more accurate distance estimates than when performing the same task in total darkness [11]. Visible environments contain visual information that houses various distance-related information sources (both monocular and binocular) like relative size, angular declination, perspective, motion parallax, binocular disparity, and convergence [62, 63], all of which can aid in perceiving depth even if specified aurally. If visual contexts can affect how far a real-world aural stimulus is perceived to be as shown

979-8-3503-7402-5/24/$31.00 ©2024 IEEE
DOI 10.1109/VR58804.2024.00089

in [11, 68], it is essential to build an understanding of how a virtual environment's visibility influences our ability to accurately gauge distances through sound as this can deepen our comprehension of human perception. Inspired by this aim and those aforementioned, we attempt to contribute to the knowledge base on the effects of calibration of egocentrically perceived aural depth under different levels of the environment's visibility.

2 Related Work and Background

2.1 Measuring Distance Perception

Egocentric estimates of distance in the medium-field are commonly measured using verbal reports, relative measures, or absolute measures [16, 21, 38, 64]. With verbal reports, users verbally estimate distance using units they are familiar with. Relative measures require users to make depth estimates based on perceptual comparisons with targets that are considered to be representative of how far a stimulus is perceived to be [45]. Researchers refer to relative measurement methods as perceptual matching tasks because estimates are obtained by having users adjust the position of a target to perceptually match the distance of a referent object, as in [39, 48, 76]. Researchers have even used perceptual comparison or forced choice tasks, having users indicate which of two or more stimuli they perceive as farther [35, 40, 60, 61]. Recent work investigating auditory depth perception used a perceptual matching task by having users indicate how far they perceived an audio source to be by selecting from a series of consecutive light bulbs placed at discrete depths [39]. These kinds of measures are often employed for their convenience or due to limited experimental space. However, these methods can introduce confounding factors, be affected by cognitive abilities, and result in less accuracy [45, 51]. In contrast to measures involving estimates based on relativity, absolute measures typically obtain egocentrically perceived distance estimates through actions that feature physical interaction between the virtual environment and the user. Examples of such measures include reach estimates made using the hand in the near-field [20], and blind-walking to the perceived location of the stimulus in the medium field. In blind-walking, participants view a target after which they are blindfolded and asked to walk to the perceived location of the target. The perceived distance to the target is determined by the distance from the participant's initial starting point to their final stopping spot. Several studies have demonstrated blind-walking's accuracy in measuring distance perception in VR. Along these lines, Loomis et al. found that blind-walking accurately measures egocentrically perceived distance in VR [41, 43]. Mohler et al. also discovered that blind-walking performance is a reliable measure of distance perception in VR, and can be employed to evaluate the effects of various visual and auditory information on perception [50]. Furthermore, it has been revealed that using auditory beeps during a blind-walking task enhanced distance perception in a VR environment [3]. Due to their reliability, absolute measures like blind-walking are considered more reliable and accurate than non-action-based relative measures [3, 13, 37, 44, 51], because estimates provided using these types of measures take into consideration the embodied action capabilities of the organism that is perceiving the environment [51].

2.2 Auditory Distance Perception and Calibration

At its core, auditory distance perception constitutes the perception of distance based on auditory information. Researchers have studied both how aural information affects the perceived depth of a stimulus in the presence of visual information of the stimulus, as well as how depth is perceived based solely on auditory information of a stimulus. It has been shown that sound can influence how we visually perceive depth in stereoscopic images and can be used to counter depth compression effects [66]. Researchers have found that both the timing as well as the type of sound play a role in shaping distance perception [14]. Along these lines, Finnegan et al. discussed using incongruent auditory and visual stimuli to compensate for distance compression effects in VR, showing that this incongruence can improve the perception of distance [25, 26]. In the presence of non-coherent audio and visual stimuli when the latter is closer than the former, it has been found that users more largely default to perceiving distances visually [53]. In an investigation of egocentric distance perception based on visual, auditory, audio-visual information, it has also been shown that distance perception is modality-independent with close distances being overestimated and farther distances being underestimated [58]. In terms of the perception of distance based on auditory information without visual information of the targets, it has been shown that distances in medium and far-field are systematically underestimated when perceived aurally [42]. Similar results were obtained by [77], showing that listeners tend to drastically underestimate the distance of an aural stimulus over longer distances. Recent work conducted in an IVE has shown that aurally perceived distance can be improved with feedback provided in a calibration phase [39]. The authors of this work discuss auditory depth perception in the context of calibration, an important concept associated with perception.

Perceptuo-motor calibration is the process of learning through tasks where participants' actions are adjusted based on corrective feedback [15, 59, 69]. Studies indicate that calibration to perceptual information can happen rapidly when individuals engage in interactive tasks with their surroundings [1, 18]. Research has demonstrated that calibration or perceptual attunement can enhance users' perceptual judgments and performance in tasks [5–7, 56, 70]. Closed-loop feedback is a widely used approach for readjusting perceptuo-motor judgments and performance in both the real world and VR, as demonstrated in studies by [17, 39]. In this method, users are provided with feedback that helps correctly fine-tune perceptual judgments, particularly when initial judgments lack accuracy due to a lack of training. Various forms of closed-loop feedback have been utilized to enhance the accuracy of estimating distances while walking in VR [50]. Notably, providing visual feedback before and after a blindfolded walk tended to enhance the precision of future blindfolded walks. Similarly, feedback was found to improve depth perception estimates obtained using physical reaches [19]. Open-loop calibration, one that does not provide feedback, led to an overestimation of distances when reaching, but closed-loop calibration significantly improved participants' ability to estimate distances more accurately [18]. More recently, researchers have shown that calibration can improve the accuracy of aurally perceived distance [39]. In this study, users heard an aural stimulus based on which they provided a depth estimate for a number of trials in a pre-test, calibration, and posttest paradigm. Feedback provided during the calibration phase was found to improve accuracy in the posttest phase. A series of light bulbs were placed on the floor of a virtual room at unit-incremental distances. Users had to light up the bulb that most accurately represented where they perceived the auditory stimulus to emanate from. This method of measurement is based on a relative estimate rather than one that is absolute, presenting an opportunity for the investigation of calibration with an absolute measure of egocentrically perceived distance, one that our work attempts to address.

2.3 Visibility of the Environment

Visual scenes contain various distance-related information sources (both monocular and binocular) like relative size, angular declination, perspective, motion parallax, binocular disparity, and convergence [62, 63], all of which can aid in perceiving depth. In VR environments, it has been shown that distances are perceived more accurately in cluttered settings because there are more visual references to base depth estimates on [46].

Lighting is an important aspect that determines an environment's visibility, making it pertinent to study how said aspect affects depth perception. Research on this front indicates that in well-lit settings, visually perceived distance is generally accurate for objects up to

20 meters away [28, 65]. In contrast, settings with less ambient illumination have been shown to degrade distance-perception accuracy [32, 55]. This implies that the visibility of the environmental context within which an object's distance is perceived can influence how far the object is visually perceived to be [22]. Recent studies have corroborated this finding, extending it to distances perceived aurally as well. Along these lines, it has been shown that affording visual information of the whole scene increases the accuracy of distance estimates made based on auditory stimuli even when the sound source is invisible [11]. This study's results showed that distance estimates made based on auditory stimuli were significantly more accurate when participants had the opportunity to observe the experimental room for a few minutes before carrying out the experiment in total darkness. These results imply that the perceived distance of the sound source was influenced by visual information about the experimental room. This finding aligns with the postulation that an environment's visual information can serve as a structured spatial reference into which distance-related information is integrated [29, 30]. The visual information of an environment tends to provide a reference frame that scales distances aurally perceived within it, expanding or constraining users' estimates. As such, there are a number of research efforts, showing that visual contexts affect aurally perceived distances through mechanisms related to multisensory integration [11, 68].

Considering the ecological approach to perception, individuals pick up invariant environmental information to perceive depth, size, and scale [7, 31, 72]. One's height is one invariant that specifies perceived depth as it is known that organisms perceive egocentric distances intrinsically [9]. If reduced visibility of an environment decreases the potential to detect visual invariants that specify the distance a sound source, then it can be expected that poorly lit rooms degrade depth perception when aural information from the sound source is, in itself, not sufficient for veridical perception. It hence behooves researchers to study how the environment's visibility affects auditory depth perception towards ascertaining if visual contextual information contains invariants that specify depth. We attempt to contribute to this knowledge base, aiming to gain insights into how an environment's visibility affects distances perceived aurally in virtual reality.

3 System Description

An HTC Vive Pro HMD equipped with noise-canceling headphones was used. Participants' movements were tracked using four HTC lighthouses arranged in a diagonal configuration. The simulation was run on a system with a Windows 11 operating system, a 2.3GHz Intel i7 processor, 32GB of RAM, and an NVIDIA GeForce RTX 3070 GPU. To minimize potential interference from environmental sounds, all experiments were conducted in a noise-controlled laboratory setting. During pilot testing, the frame-rate of the simulation was measured on multiple occasions, ensuring that it was stable and approximately equal to the device's maximum refresh rate (90Hz).

3.1 Virtual Scenes

We created two virtual scenes using the Unreal 4.27 game engine: a dark room and a light room (Fig. 1). The light room had a centrally positioned rectangular light on the ceiling, illuminating the entire room with an intensity of 120 cd. Both scenes had the same dimensions of 8.72m × 3.5m × 15m. The starting location was marked by a yellow circle (diameter = 1m) on the floor. A red circle (diameter = 1m) indicated the participants' perceived location of the target, and a green circle (diameter = 1m) indicated the actual target's location. The walls, ceiling, and floor of the room did not have any periodic patterns or distinctive textures to avoid participants relying on them for additional visual information when making depth judgments.

3.2 Audio Stimuli

Our VR system leverages the power of binaural rendering, which is fundamental to creating an immersive VR experience by enhancing the perception of three-dimensional space and increasing the sense of presence in the virtual environment [12, 23]. As corroborated by previous research, the use of spatial audio can notably enhance presence, immersion, engagement, and performance in VR applications such as gaming and training simulations [36, 52]. We intentionally selected a four-second-long horn sound from an anechoic orchestra recording for our study [54]. This choice of a non-melodic segment was deliberately chosen to avoid using audio stimuli that participants may be familiar with, as used in [39]. This is because familiarity with sounds can assist participants in estimating distance, as indicated in previous research [49].

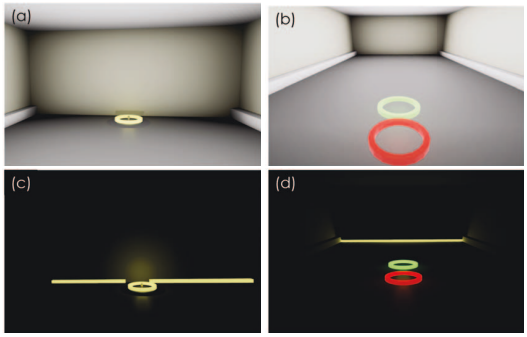

Figure 1: Virtual scenes in the experiment: (a)(b) more visible and (c)(d) less visible environments. Yellow circles in (a) and (c) indicate the starting location; Red circles in (b) and (d) represent a user's perceived (and hence judged) target location; Green circles in (b) and (d) represent the target's actual location. Subfigures (b) and (d) correspond to the feedback stage during the calibration phase

We used Steam Audio for spatial audio rendering of the anechoic horn sound. It simulates physics-based sound behaviors in VEs [4] using ray tracing to perform the necessary auralization of the designed room in real time, including aspects of sound spatialization, propagation, reflection, occlusion, transmission, and reverberation. It is a powerful sound engine adopted in recent studies [4, 10]. Steam Audio's physics-based sound behavior is a critical feature of our system, facilitating real-time sound spatialization and propagation with high-quality audio responses that can be customized for various scenarios and tasks. All the audio stimuli were rendered with the default settings [67] of Steam Audio and the room geometry and locations of the sound source were specified according to the designed virtual scenes. Our system incorporates virtual spherical speakers that are placed on the ground at the set targets' locations (Table 1). These speakers were made invisible within the virtual scenes. To eliminate the possibility of participants memorizing and counting steps for each actual distance, we divided the target distances into two sets, using one in each of the phases described in section 4.1. The target distances used in the different phases are listed in Table 1, and pilot tests confirmed that all stimuli from these distances were audible but not uncomfortably loud.

Pretest/Posttest	3m	5m	7m	9m	11m
Calibration	2m	4m	6m	8m	10m

Table 1: Set targets' distances of audio stimuli in the three phases.

4 Experiment

4.1 Study Design

To empirically study the calibration of aurally perceived depth (measured using an absolute action-based measure) under different levels of a virtual environment's visibility, we employed a 2 (Visibility of virtual environment) × 3 (Phase) × 5 (Target Distance) multifactorial design, manipulating the phase and target distance as within-subjects factors, and the visibility of the virtual environment as a

between-subjects factor. Two levels of environment visibility were studied: (1) Less visible (LV); (2) More visible (MV). In the more visible condition, the room was lit brightly and was completely visible to the user. In the less visible environment, rather than complete darkness, a strip of volumetric lights was added to the far wall of the room, allowing users to see the extent of the room and the floor (Fig. 1). Participants were randomly assigned to either of these two levels where they performed an auditory information-based egocentric absolute depth judgment task (section 4.2) in 3 sequential phases: pretest, calibration, and posttest. In the pretest phase, participants made action-based depth judgments of how far they perceived a target to be without any feedback about their judgment accuracy. In the calibration phase, however, participants were provided with multimodal feedback based on which they could adjust or rather calibrate their depth judgments. This provided users with the opportunity to observe the offset between the original location of the target stimulus (sound source) and the location at which they perceived and hence judged this stimulus to be. Similar to methods employed in [20,39], the calibration phase allowed for users to make corrective actions toward improving their depth judgment accuracy. The posttest phase was identical to the pretest phase, thus allowing us to study the effect of calibration and perceptual learning. In each phase, the trials featured the targets placed at five different distances (Table 1) each of which was repeated five times, thus accruing up to a total of 25 trials. To prevent order effects, the order of these 25 trials was randomized. With every participant undergoing all three phases, this study involved each user making auditory depth judgments for a total of 75 trials.

4.2 Tasks

A simple auditory-information-based absolute depth judgment task was conceptualized wherein participants had to judge how far an invisible sound source was located from them, for a number of trials in VR. The stages involved in the task depended on the phase in which participants performed it. In the pre and posttest phases, participants performed open-loop distance estimation without feedback of their judgment accuracy. In contrast, users performed closed-loop distance estimation with feedback in the calibration phase.

4.2.1 Open-loop Task in Pretest and Posttest Phases

In both of these phases, each trial consisted of two sequential stages: perception and response (Fig. 2). In the perception stage, participants stood at a fixed starting location that was denoted by a yellow circle. They would then hear an auditory stimulus emanating from an invisible sound source (target) which was located at a specific distance (target distance) from them. Once they indicated that they were ready, the sound stimulus was stopped. This marked the completion of the perception stage and start of the response stage. In this response stage, the participants would *silent-walk* (in the absence of the auditory stimulus) to the location at which they perceived the sound to emanate from. They would then confirm their depth judgment by signalling to the experimenter. This marked the completion of the response stage and therefore, the trial. Users then walked back to the starting location to initiate the next trial. This *silent-walking* technique can be considered an auditory analog to blind-walking, an absolute measurement technique commonly used in medium-field depth estimation studies to measure visually perceived distances.

4.2.2 Closed-loop Task in Calibration Phase

In this phase, each trial consisted of the four sequential stages: perception, response, feedback, and correction; this is adopted from perception-action based calibration literature [20, 59]. Fig. 2 depicts the protocol adopted for the calibration phase. The perception (steps 1, 2 and 3) and response (steps 5 and 6) stages in these trials were identical to the pretest and posttest phases. Unlike those phases, however, the completion of the response stage was immediately followed by a feedback stage. In this stage (steps 8, 9, and 10), participants were provided with multimodal feedback of their depth judgment accuracy after being teleported back to the starting location from where they could once again hear the sound stimulus (as originally heard in the perception phase), see a visual depiction of the target's actual location (denoted by a green circle), and see a visual depiction of the location at which they judged that target to be (denoted by a red circle)(Fig. 1). After observing the disparity between their judgment of the target's location and its actual location, participants would signal to the experimenter, marking the completion of the feedback stage. They were then teleported back to their judgment location (step 7) after which the correction stage commenced. This stage (steps 12 and 13 in Fig. 2) involved participants making a corrective walk to the target's actual location (green circle) in the presence of the audio stimulus. Upon reaching the target's actual location, users would signal to the experimenter, marking the completion of the correction stage. The sound stimulus was immediately stopped, following which participants would then walk back to the starting location to initiate the next trial (steps 14 and 15). It was ensured that the five target distances presented in the calibration phase were different from those tested in the pretest and posttest phase to prevent learning effects that could have compromised judgments made in the posttest phase.

The audio stimulus used in this study was always spatially rendered as a function of the distance between the participant and the sound source, taking into account the environment's properties. Thus, the sound stimulus heard by users during the perception and feedback stages were identical given that they were standing at the starting location in both of these stages. The sound stimulus heard by users at the start of the correction stage accurately represented how it would be to hear that sound from the judgment location. During the corrective walk, the sound would naturally intensify as the user approached the actual location of the target.

The rationale for teleporting users back to the starting location during the feedback stage was to precisely control and re-present the original auditory stimulus heard during the perception stage. If users were not teleported back to the starting location, the sound heard would not match the original sound stimulus presented in the perception stage simply by virtue of the user's position at that instant (judgment location). Teleporting users back to the starting location during the feedback stage hence ensured that feedback was provided in a manner that gave visual indications of their performance (disparity between red and green circles) in the presence of the original stimulus. Fig. 2 illustrates a schematic representation of the stages and events that transpired in each trial of the calibration phase. Since users were teleported in VR during the different stages of calibration, the final walking distance including the corrective walk amounted to exactly the total distance of the stimuli (except in overestimation trials). However, it should be noted that our protocol required users to walk for every trial, as in the case of blind-walking.

Based on users' judgements, each trial resulted in one of three cases: (1) Underestimation: when the judged location of target was closer than its actual location (red circle is closer than the green circle); (2) Overestimation: when the judged location of target was further than its actual location (red circle is further beyond the green circle); (3) Correct estimation: when the judged location of the target coincided with its actual location (red and green circles overlap).

4.3 Research Questions and Hypotheses

The overarching aim of this work was to answer the following research question: *can auditory depth perception be calibrated using an absolute measure of distance estimation?* Downstream of this, we were interested in understanding how the environment's visibility affects users' auditory depth perception accuracy. We developed the following hypotheses based on work discussed in section 2:

H1 (Absolute measure of auditory depth perception): The target's actual location will significantly predict the perceived location of the target with high levels of goodness of fit, suggesting validity in

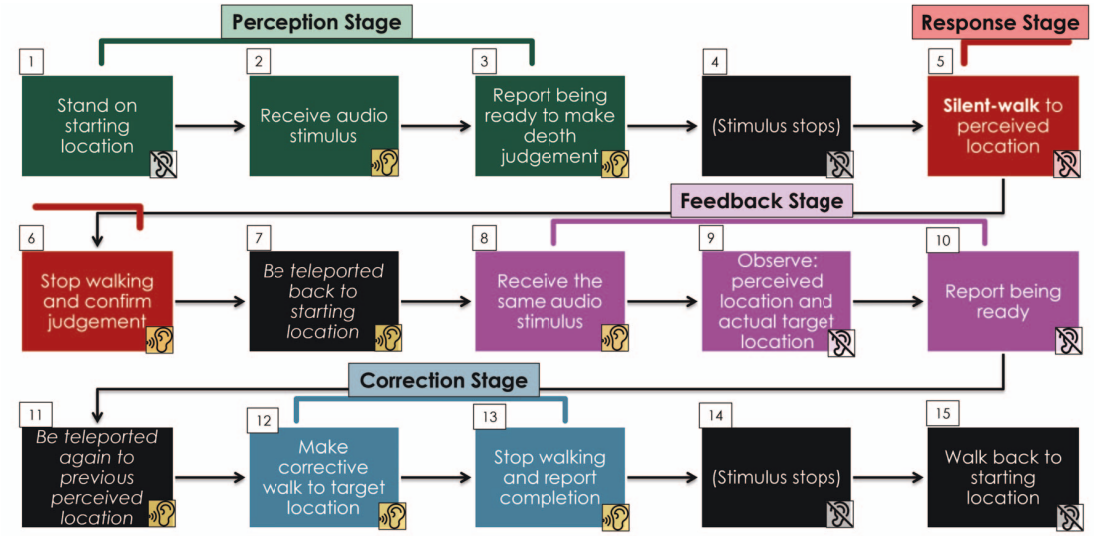

Figure 2: **Open-loop task** in the pretest/posttest phase only consists of steps 1, 2, 3, 4, 5, 6, and 15, in which participants perceive stimuli without feedback. **Closed-loop task** in the calibration phase consists of all the steps that participants perceive with feedback by teleportation back to starting location. The ear icons located in the lower right corner signify the participant's current status regarding auditory stimuli. The steps are color coded based on the stages occurring in the calibration phase.

the silent-walking method to measure auditory depth perception.

H2 (Calibration of auditory depth perception): Participants' performance in the posttest phase will be more accurate and less erroneous than in the pretest phase, indicating the efficacy of the calibration phase in facilitating perceptual learning.

H3 (Effects of environment's visibility): The more visible environment will produce higher levels of depth perception accuracy compared to the low visible environment.

Based on studies that have demonstrated the efficacy of blind-walking as an action-based absolute measure of egocentric visual depth perception in VR, it stands to reason that its analog, *silent-walking*, should be equally effective in serving as a measure of auditory depth perception. For this reason, it is expected that for the task described in section 4.2, the actual target's distance should be a robust predictor of users' perceived distance in the trials.

Perceptual learning has been extensively studied and shown to exist in various sensory modalities in IVEs, e.g., visuomotor calibration [19, 70] and visuo-auditory distance perception [39]. Along these lines, calibration to auditory reverberation time has been shown to exist in auditory distance perception, improving participants' depth judgment accuracy [39]. Given that such learning effects exist in auditory depth perception, it is expected that calibration occurs when using an action-based absolute measure like *silent-walking*. For this reason, one can anticipate that participants' performance in the posttest phase will be more accurate and less erroneous than in the pretest phase. Environmental properties affect how auditory information is perceived. With the audio in this study being spatially modeled and rendered, taking into account the environment's properties (dimensions of the room, materials on wall, floor, ceiling, etc.), it is expected that a more visible environment should improve depth perception accuracy because, in addition to the auditory information, users in such an environment have more visual information of the environment which, in part, determines how sound is heard.

4.4 Participants

To determine an appropriate sample size for our experiment, we conducted an apriori power analysis using G*Power software. Based on an effect size of 0.25 from a previous study [20], an alpha error probability of 0.05, power of 0.95, 25 measurements per session, and a correlation of 0.5 among repeated measures, the analysis indicated that we needed at least 12 participants in one condition. We ultimately recruited a total of 30 participants for this Institution Review Board approved study with 8 females and 7 males in the more visible condition and 10 females and 5 males in the less visible condition. Prior to the experiment, participants were assessed for their hearing ability using personal earphones and the Widex hearing evaluation website [73] to make sure they met the inclusion criteria of normal hearing in both ears, 20/20 visual acuity or corrected vision, and the ability to fuse stereoscopic images. Participants' ages ranged from 21 to 30 years old (M=23.3, STD=3.62) , 29 of whom were students. All participants had gaming experiences and only six of them had never experienced VR previously. Overall, VR experience did not significantly across conditions.

4.5 Procedure

Upon arrival, participants were greeted and asked to read and sign a consent form. They then filled out a demographics survey including information about their backgrounds along with their VR and gaming experience. Following this, participants' interpupillary distances (IPD) were measured using a digital pupil distance meter, and this was adjusted in the HMD. Participants were then instructed to stand on a fixed point for accurate positioning, and their eye height was adjusted to correspond to their real-world height [24]. Users were then randomly assigned to one of the two visibility conditions after which the experimenter briefed them about the task that they would be performing in the experiment. Following this, participants performed three practice trials to familiarize themselves with the task and its specifics. It was ensured that the distances presented in the practice trials were different from those used in the experiment, thus avoiding any potential effects of learning. Upon completion of the practice trials, participants began the experiment performing the trials across the three sequential phases (section 4.2). After completing all three phases, participants filled out the NASA TLX workload questionnaire. Upon completion, participants were debriefed and compensated for their time. On average, it took participants up to two hours to complete the study.

4.6 Measures

Perceived distance (PD) - It represents how far participants egocentrically perceived the sound source to be from the starting location. In other words, PD is the participant's response in a distance estimation task, reflecting their perceived distance to the target location.

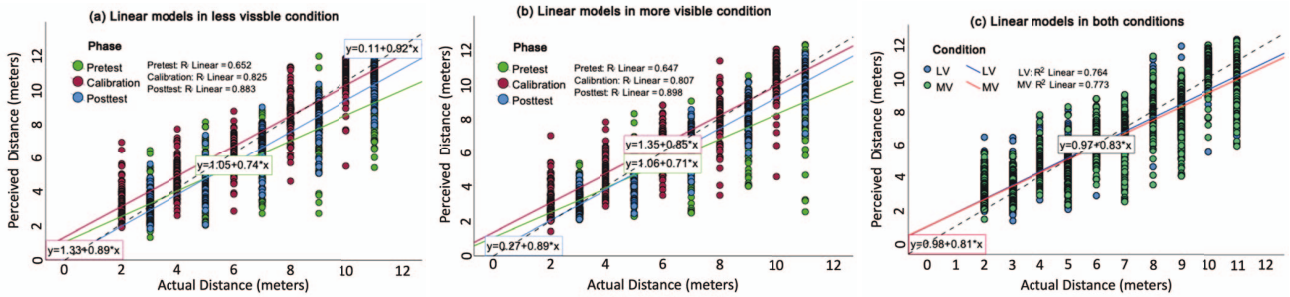

Figure 3: Linear regression models showing the prediction of perceived distance based on the sound source's actual distance across the pretest, calibration, and posttest phases in (a) less visible environment, (b) more visible environment;(c) Regression model averaged across all phases in the two levels of the environment's visibility.

Signed Error - The accuracy of participants' distance judgments was operationalized as the signed error computed as the difference between a participants' perceived distance of the target and the actual target's distance from the start location. Mathematically, *Signed error = Perceived Distance - Actual Distance*. Naturally, a positive *signed error* indicates an overestimation of distance, while a negative one indicates an underestimation. For each trial in every phase, we calculated the *signed error* across both the visibility conditions.

Stimuli Time (ST) - The stimuli time represents the amount of time that participants required the audio stimulus before they were ready to make their depth judgment. It denotes the time interval between the moment a participant began to hear the audio stimuli and the moment they signaled to stop the audio. A longer stimuli time implies that participants needed more time before they were ready to make their distance judgment. Mathematically, it was calculated as the difference in timestamps between the start and the stop (signalled by participants) of the audio stimulus.

Workload (NASA-TLX) - Users' perceived level of workload due to the simulation was measured using NASA TLX questionnaire [33].

5 Results

5.1 Multiple Regression Results

To evaluate how phase of measurement of the depth perception (pretest/calibration/posttest), actual distance, condition, condition-by-actual distance interaction, phase-by-condition interaction, phase-by-actual distance interaction, and three-way condition-by-phase-by-actual distance interaction affect the perceived distance to the targets, a multiple regression was calculated to predict perceived distance based on the independent variables.

As is often done ahead of a multiple regression analysis, an analysis of standardized residuals was carried out on the data to identify any outliers, using which data that were beyond $\pm$ 3.0 of the standardized residuals were removed. The final standardized residual minimum was -1.50 to +1.50. Tests to examine if the data met the assumptions of collinearity indicated that multi-collinearity was not a concern (Phase, Tolerance = 1.0, VIF = 1.0; Actual Distance, Tolerance = 1.0, VIF = 1.0; Condition, Tolerance = 1.0, VIF = 1.0). The data met the assumptions of independent errors (Durbin-Watson value = 1.31). The histogram of standardized residuals indicated that the data contained approximately normally distributed errors, as did the P-P plot of standardized residuals, which showed that the data were close to a linear regression profile. The scatter-plot of standardized residuals showed that the data met the assumptions of homogeneity of variance and linearity, as well as the data met the assumptions of non-zero variance.

A multiple regression was first conducted to examine if phase, condition, and actual distance predicted perceived distance. A significant regression equation was found $F(3, 2237) = 2502.53$, $p < 0.001$, with an $R^2 = 0.77$. Participants' *Perceived Distance = 0.83 + 0.82 $\times$ Actual Distance (partial R^2 = 0.77) - 0.12 $\times$ Condition (partial R^2 = 0.0025) + 0.17$\times$ Phase (partial R^2 = 0.01)*; where *Actual Distance* was measured in meters, phase 1 was pretest, 2 was calibration, 3 was posttest; condition 1 was less visible condition (LV), condition 2 was more visible condition (MV). *Perceived Distance* increased by 0.82 meters for every meter of increase in actual distance, perceived distance decreased by 0.12 meters for a change in condition, perceived distance increased by 0.17 meters for a difference in one phase to the next (i.e., pretest to calibration or calibration to posttest). All of the three factors: *Actual Distance* ($p < 0.001$) and *Phase* ($p < 0.001$) and *Condition* ($p < 0.001$) were significant predictors of *Perceived Distance*.

In order to evaluate the significant interaction effects, the continuous independent variable of actual distance was mean centered to eliminate any multicollinearity effects, and the interaction terms (mean centered) *Actual Distance $\times$ Condition*, *Actual Distance $\times$ Phase*, *Phase $\times$ Condition*, and *Actual Distance $\times$ Condition $\times$ Phase* were added to the model in a hierarchical multiple regression. The regression model with the interaction variables was found to be significant, $F(7, 2237) = 1099.63$, $p < 0.001$, with an $R^2 = 0.78$ (with the change in R^2 of 0.005). When including the interaction term, participant's *Perceived Distance = 2.069 + 0.666 $\times$ Actual Distance – 0.239 $\times$ Condition + 0.055 $\times$ Phase – 0.002 (Actual Distance $\times$ Condition) + 0.091 (Actual Distance $\times$ Phase) + 0.057 (Condition $\times$ Phase) – 0.009 (Actual Distance $\times$ Condition $\times$ Phase)* ; where *Actual Distance* was measured in meters, phase 1 was pretest, 2 was calibration, 3 was posttest; condition 1 was LV, condition 2 was MV. *Perceived Distance* increased by 0.666 meters for every meter of increase in actual distance, perceived distance decreased by 0.239 meters for a changing in *Condition*, and perceived distance increased by 0.055 meters for a difference in one *Phase* to the next (i.e., pretest to calibration or calibration to posttest), perceived distance decreased 0.002 for a unit of *Condition* by *Actual Distance* interaction, perceived distance increased 0.091 for a unit of *Phase* by *Actual Distance* interaction, perceived distance increased 0.009 for a unit of *Condition* by *Phase* by *Actual Distance* interaction. All seven predictors, *Actual Distance* ($p < 0.001$), *Condition* ($p < 0.001$), *Phase* ($p < 0.001$), *Condition* by *Actual Distance* interaction term ($p < 0.001$), *Phase* by *Actual Distance* interaction term ($p < 0.001$), *Condition* by *Phase* interaction term ($p < 0.001$), and *Actual Distance* by *Phase* by *Actual Distance* interaction term ($p < 0.001$), were significant predictors of *Perceived Distance*.

By phase, the linear regression equation in less visible condition for the pretest phase (R^2=0.65) is *Perceived Distance = 1.05 + 0.74 $\times$ Actual Distance*, the linear regression equation for the initial judgments in the calibration phase is (R^2=0.83) is *Perceived Distance = 1.33 + 0.89 $\times$ Actual Distance*, the linear regression equation for the posttest phase is (R^2=0.88) is *Perceived Distance = 0.11 + 0.92 $\times$ Actual Distance* (Fig. 3a).

By phase, the linear regression equation in more visible condition for the pretest phase (R^2=0.65) is *Perceived Distance = 1.06 + 0.71 $\times$ Actual Distance*, the linear regression equation for the initial

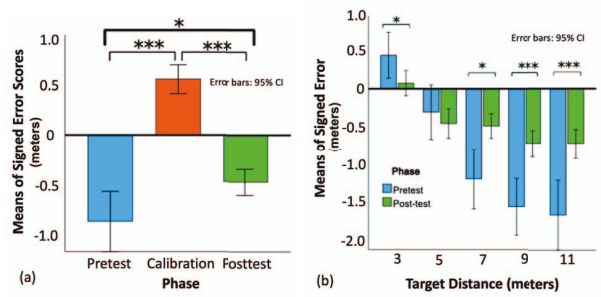

Figure 4: Mean signed error scores by (a) phases, and (b) target distances.

judgments in the calibration phase is (R^2=0.81) is *Perceived Distance = 1.35 + 0.85 × Actual Distance*, the linear regression equation for the posttest phase is (R^2=0.90) is *Perceived Distance = 0.27 + 0.89 × Actual Distance* (Fig. 3b).

5.2 Depth Perception Accuracy

To analyze participants' depth perception accuracy across the different conditions and phases, we applied a 3×2 mixed model ANOVA, investigating the effects of phase (within-subjects factor) and condition (between-subjects factor) on signed errors after carefully verifying that the assumptions of the mixed ANOVA were met. The data in the samples were normally distributed and error variance between signed error scores in the different phases was equivalent. We insured that Box's test of equality of covariance matrix was not significant. Levene's test was conducted to verify homogeneity of variance, and Mauchly's test of sphericity was conducted to ensure that error variance between *signed error scores* in the different phases of the experiment was equivalent. Pairwise post-hoc tests were conducted using Bonferroni adjusted alpha method. We averaged the signed error generated in three phases and two different conditions for this analysis.

The ANOVA analysis revealed a significant main effect of phase F(2,56) = 71.31, p <0.001, partial.η^2 = 0.72. Post-hoc pairwise comparisons using Bonferroni method revealed that participants' signed error scores were significantly higher in the calibration phase (M=0.56, SD=0.39) as compared to their initial judgment in the pretest (M=−0.86, SD=0.79) p <0.001, and in the posttest phase (M=−0.47, SD=0.34) p <0.001. Participants' signed error scores were also significantly higher in the posttest phase (M=−0.47, SD=0.34) as compared to pretest (M=−0.86, SD=0.79) p <0.005. Mean signed error scores in calibration phase was the highest (Fig. 4a). We didn't find a main effect of condition nor were any other interaction effects significant.

We conducted another analysis of signed error by different target distances, comparing the signed error by target between pretest and posttest phases. The targets used in the experiment were located at 3, 5, 7, 9, and 11 meters, which were the same for both phases. Before performing a parametric repeated measures ANOVA analysis on the signed error by target scores, we confirmed that the underlying assumptions of the test were met. Specifically, we checked for normal distribution of data in the samples and equivalent error variance between signed error by target in the different phases and conditions. We also ensured that Box's test of equality of covariance matrix was not significant and conducted Levene's test to verify homogeneity of variance. Mauchly's test of sphericity was performed to ensure that error variance between signed error by target scores in different phases of the experiment was equal. Finally, pairwise post-hoc tests were conducted using the Bonferroni adjusted alpha method.

The ANOVA analysis revealed another significant main effect of target distance F(4,112) = 72.16, p <0.001, partial.η^2 = 0.72. Post-hoc pairwise comparisons using Bonferroni method revealed that participants' signed error by target distances were significantly higher when the target was located at 3 meters (M=0.26, SD=0.10) as compared to when the target was located at 5 meters (M=−0.38, SD=0.11), p <0.001; 7 meters (M=−0.85, SD=0.10), p <0.001; 9 meters (M=−1.15, SD=0.10), p <0.001; 11 meters (M=−1.20, SD=0.12),p <0.001; participants' accuracy when the target distance was 5 meters was also significantly higher as compared to targets located at 7 meters, 9 meters, and 11 meters.

The ANOVA analysis revealed another significant phase by target interaction effect F(4,112) = 16.69, p <0.01, partial.η^2 = 0.37. As shown in Fig. 4b, post-hoc pairwise comparisons using Bonferroni method revealed that participants' signed error by target distances were significantly higher when the target was located at 3 meters in pretest (M=0.442, SD=0.15) as compared to posttest (M=0.07, SD=0.08),p <0.05; 7 meters in pretest (M=−1.20, SD=0.19) as compared to posttest (M=−0.49, SD=0.08), p <0.05; 9 meters in pretest (M=−1.57, SD=0.18) as compared to posttest (M=−0.73, SD=0.08), p <0.001; and 11 meters in the pretest (M=−1.67, SD=0.23) as compared to in the posttest (M=−0.73, SD=0.09), p <0.001. No other main or interaction effects were found.

5.3 Stimuli Time

The stimuli time data was carefully verified to insure that the underlying assumption of a parametric analysis was met. The data was normally distributed and variance was homogeneous among groups of stimuli time scores between the experiment phases and conditions. The mean stimuli time between experiment phases and conditions were subjected to a 3 (phase) × 2 (condition) mixed model ANOVA analysis. Post-hoc comparisons on the within-subjects factors were conducted using the Bonferroni method.

Prior to conducting the parametric mixed model ANOVA analysis on the stimuli time, we carefully verified that the underlying assumptions of the test were met. Namely, the data in the samples were normally distributed and error variance between stimuli time in different phases were equivalent. We insured that Box's test of equality of covariance matrix was not significant. Levene's test was conducted to verify homogeneity of variance, and Mauchly's test of sphericity was conducted to ensure that error variance between stimuli time in the different phases was equivalent. Pairwise post-hoc tests were conducted using Bonferroni adjusted alpha method.

We firstly calculate the overall ST to a 3 (phase) × 2 (condition) mixed model ANOVA analysis. Phase was a within-subjects variable consisting of pretest, calibration, posttest; condition was a between-subjects variable consisting of less and more visible conditions. The ANOVA analysis revealed a significant main effect of condition, F(1,28) = 5.12, p < 0.05, partial.η^2 = 0.16. The ANOVA analysis also revealed a significant main effect of phase, F(2,56) = 34.18, p < 0.001, partial.η^2 = 0.55. Post-hoc pairwise comparisons using Bonferroni method revealed that participants' mean stimuli time were significantly longer in the pretest phase (M=9.98, SD=0.74) as compared to the calibration phase (M=7.16, SD=2.69) p <0.001, and to the posttest phase (M=6.46, SD=2.84) p <0.001; mean stimuli time in pretest phase was the longest. The ANOVA analysis revealed another significant main effect of phase by condition, F(2,56) = 3.33, p < 0.05, partial.η^2 = 0.10. Post-hoc pairwise comparisons using Bonferroni method revealed that participants' mean stimuli time in the posttest phase time were significantly longer in less visible condition (M=7.57, SD=3.06) as compared to in more visible condition (M=5.34, SD=2.16) p < 0.05 (Fig. 5b).

5.4 NASA-TLX

A Mann Whitney U test on Temporal demand scores revealed that the less visible (LV) condition (M=4.53, SD=5.33) was significantly higher than the more visible (MV) condition (M=0.53, SD=1.36), U = 45, p < 0.05; Analysis on effort scores in the pretest revealed that the LV condition (M=21.73, SD=9.13) was significantly higher than the MV condition (M=15.47, SD=9.48), U = 63.5, p < 0.05;

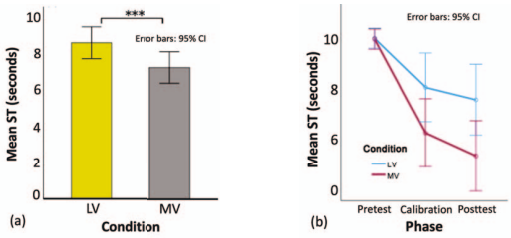

Figure 5: Mean stimuli time taken in a) conditions and b) phases broken down by condition.

Analysis on frustration scores revealed that the LV condition (M=7.6, SD=9.4) was significantly lower than the MV condition (M=11.47, SD=11.39), $U = 172$, $p < 0.05$.

6 Discussion

Silent Walking - The regression analyses revealed the existence of a significant relationship between the perceived distance of a target and its actual distance. The regression model shows that as the actual distance of the target increases, users' perception of the distance to that target also increases (Fig. 3). There is a linear relationship between the actual distance and participants' perceived distances of the sound source, featuring a slope that is close to 1 and an intercept that is close to 0. Additionally, a large proportion of variance in the perceived distance (77%) is explained by the regression model, suggesting that participants are highly consistent as well. With the actual distance being a highly significant predictor of perceived distance, explaining a large proportion in its variance, we obtained support for hypothesis H1, in that *silent-walking* seems to be a valid and reliable method of measuring auditory depth perception using an absolute measure as in the case of blind walking for visual depth perception. This method can serve as a useful tool for researchers studying this phenomenon.

Calibration of Auditory Depth Perception - On inspecting the regression models in relation to learning effects, we found that participants were able to calibrate their perception of depth to the spatial auditory information, producing more accurate (less erroneous) estimates of egocentric distances after enduring the calibration phase. This finding can be inferred from Figs. 3a and 3b which depicts the effects of the actual distance on perceived distance moderated by the experimental phase. The ideal prediction is represented by the dotted line which characterizes the veridical judgment, the slope of which equals one. The veridical judgment line represents the prediction where participants' perceived distance is exactly equal to the actual distance of the presented stimulus, representing 100% accuracy. A regression line that is closer to the veridical judgment line implies that the perceived egocentric distances are more accurate. As can be seen in the aforementioned figures, the regression lines of the pretest phase have a shallower slope and are less close to the veridical line than the lines of the posttest phase. This means that the degree of depth underestimation was greatest in the pretest phase and lowest in the posttest phase, implying that perceptual accuracy was higher in the latter. Identical trends were observed in terms of the signed errors in the posttest and pretest phases (Fig. 4) . As can be seen in this figure, users' estimates were underestimated to a smaller degree in the posttest phase as compared to the pretest phase, implying higher accuracy in the former. These results, along with those observed from our regression models, are indicative of the success in employing a calibration phase toward improving participants' egocentric auditory information-based distance estimation, thus supporting hypothesis H2. Our findings seem to be in line with other works that have shown that users can successfully calibrate their perception of depth using feedback obtained during a calibration phase [1, 7, 18]. Similar to the results obtained by [39], we also found that employing a calibration phase facilitated perceptual learning of auditory depth perception by improving accuracy. We extend their results, further demonstrating calibration of auditory depth perception obtained using an absolute rather than a relative measure of aurally perceived distance when using realistic spatial audio stimuli rather than exaggerated reverberation information.

Our analyses of the regression models and those conducted on the signed errors indicate the aurally perceived distances generally seem to be underestimated in IVEs. This aligns with real-world research showing that distances in medium and far-field ranges are systematically underestimated when perceived aurally [42, 77]. Similar to the depth compression effects that occur in visually perceived information, we find that users seem to systematically underestimate aurally perceived distances in IVEs as well, directly converging with findings obtained in [39]. Compared to the exaggerated reverberation used to facilitate calibration in the aforementioned study, however, our results show that users can be trained to improve accuracy using realistic spatial audio, providing a more applicable solution to improve auditory depth perception in IVEs.

Environment Visibility - With regard to the visibility of the environment and its effects on auditory depth perception accuracy, we found no significant main or interaction effect associated with condition on the signed error scores and the regression results (Fig. 3c). This suggested that the increased visibility of the virtual environment did not increase the accuracy of depth estimates and that the auditory information was what users primarily attuned to in providing their auditory perceptual estimates. We expected the increased visibility of the environment to aid participants in their auditory depth perception, improving the accuracy of their estimates, based on prior research showing that visual context information can improve auditory depth perception [11, 68]. Instead, we found that the perceptual accuracy was more or less the same in both the less and more visible conditions, thereby countering our predictions made in Hypothesis H3. It is possible that the less visible environment, though less illuminated (Fig. 1), offered sufficient visual information about the extent of the room, for participants to be accurate in perceiving the depth of the stimuli. Research on this front has shown that users' perceptions of distance to an aural stimulus heard in complete darkness become more accurate if they are shown the room prior to the experiment [11]. The exposition offered is that an environment's visual information can serve as a structured spatial reference into which distance-related information is integrated [29, 30], providing a reference frame that scales distances aurally perceived within it, expanding or constraining estimates of distance. Along these lines, the extent of the room being discernible even in the less visible environment may have afforded users the spatial reference information required to be equally as accurate in perceiving depth under poor illumination than under higher illumination. Our results lead us to two insights; firstly, compared to visual information, auditory information is a more potent determinant of how users perceive the distance of aural stimuli; and second, information about the extent of the room in which an aural stimulus is housed seems to contain information that aids in perceiving auditory depth. If we run an additional experimental condition where users provide aural depth estimates in total darkness, we would be able to ascertain whether the room's extent is the information that participants attune to in scaling their depth perception under poorly lit environments. Towards this end, future investigations must amplify the effect of visibility, investigating auditory depth perception under a truly dark IVE, one that does not afford any visual information whatsoever.

Time taken to make judgments - On the subject of how long participants required the auditory stimulus before they were ready to make depth judgments, we found a rather interesting effect of visibility of the environment. It was observed that participants required the auditory stimulus for a longer time when the environment was less visible, as compared to when the

environment was more visible (Fig. 5a). This makes sense because a more visible environment offers a greater amount of invariant information that users can attune to, decreasing the time required to make perceptual judgments. Research on this front has shown that a larger availability of invariant information yields faster response times when making perceptual judgments about axes of rotations [57]. Our results serve to extend this finding obtained in the aforementioned study, broadening its application to perceptual judgments that involve auditory depth perception. It is also possible that participants required the auditory stimuli for a longer time when the environment was less visible simply because it could have taken a long time to visually localize the location to which they had to walk to. While the LV condition was not completely dark, it did feature relatively lower levels of illumination. In contrast, the environment in the MV condition was highly illuminated, possibly allowing participants to more quickly, and visually localize the location to which they had to walk to in each trial. Interestingly, we did not find a relationship between how long participants received the auditory stimuli and how accurate they were in their depth judgments. Even though participants in the LV condition perceived the stimuli for a longer time, they were not more accurate in their depth estimates than those in the MV condition. We did, however, observe that the time required with the stimuli decreased over sessions as evinced in Fig. 5b. This is understandable with users possibly learning how to perform the task over trials.

Workload - The analyses of the NASA TLX scores indicated that participants in the less visible condition reported higher levels of temporal demand and effort compared to those in the more visible condition. This finding was consistent with participants' interview responses with several users commenting about the challenges associated with walking in the low visibility environment. The results suggest that the cognitive demand associated with perceiving the distance of an aural stimulus in poorly lit environments is more than doing the same in brightly lit environments.

7 Summary, Scope, and Limitations

Overall, this work provided us with some interesting findings. First, we discussed the usage of an action-based absolute distance estimation technique, *silent-walking*, for egocentric auditory depth perception in Virtual Reality. We found this to be a useful and valid means of measuring the perceived distance of auditory stimuli in virtual experiences, allowing researchers studying such phenomena to leverage this method of measurement. Second, we found that similar to visual depth perception, users systematically underestimate distances to aural stimuli as well. This being said, we also found that users can calibrate their spatial auditory depth perception when using an action-based absolute response measure of perceived depth estimates. It was observed that upon experiencing a calibration phase with feedback provided about depth estimation accuracy, the accuracy of their auditory depth perception increased. This adds to findings obtained in earlier work showing such effects of calibration of auditory depth estimates obtained using a relative measure, extending the same to estimates obtained from absolute measures. Third, we found that the visibility of the environment did not have a significant impact on users' auditory depth perception accuracy possibly because the auditory information contains the invariant information that users primarily attune to in making depth judgments for such stimuli. Poorly lit scenes that still reveal the environment's extent may contain information that scales distance estimates. Interestingly, we found that lower visibility of the environment increased the amount of time users required the stimuli before they were ready to make judgments. In total, we find that auditory depth perception can be calibrated under different levels of environment visibilities.

Silent-walking is an auditory analog to blind-walking, a technique commonly used to measure egocentric depth perception of visual information. Despite their similarities, the protocol of *silent-walking* requires that the users be teleported to the starting location during the feedback stage in order to ensure that the feedback provided matches the original stimulus perceived in the perception stage. Teleporting users back and forth, although necessary to avoid confounds, can cause some sickness and can hence be considered a limitation of the *silent-walking* protocol used to measure auditory depth perception.

The results of our study indicated that there were no detriments to auditory depth perception accuracy when the environment was less visible. It is possible the environment used in this study was not dark enough to elicit an effect on this front. A truly dark environment will not provide users with any visual information about the room, making for a deeper investigation of the effects of environmental visibility. Thus, despite the low level of visibility, it is plausible that the visual information perceivable in this kind of environment still provided users with information to scale their aurally perceived distances. This can be considered another limitation of this work.

Finally, the findings of this study may not be reproduced if other audio simulators are used. This is because the audio spatializer could influence participants' sound localization, and hence their auditory distance perception. However, since our study focused on auditory depth perception, in which the sound source is located in front of the participants, the influence of localization, particularly the sound direction, is minimized. Therefore, our findings could still be valid if other audio spatializers were used. This is an interesting direction for future work. Although our findings could be limited to auditory stimuli generated by Steam Audio, their applicability is sufficiently general, as Steam Audio is commonly used in VR.

8 Conclusion

In this study, we discuss the usage of *silent-walking*, an action-based absolute measure of egocentrically perceived distances of auditory information in VR, arming researchers with a robust measurement method of auditory depth perception. We observe that individuals commonly underestimate distances of sound sources in VR, much like how distances are underestimated when relying on visual perception in VR. We conducted a study, evaluating the feasibility of calibration to improve upon auditory distance perception in IVEs by provisioning closed-loop feedback on auditory depth estimates provided during a calibration phase. Our results revealed that users' auditory depth estimates can be calibrated to become more accurate with the provision of such feedback. We further evaluated the effects of the environment's visibility on auditory depth perception, finding that environments visible enough to perceive their extent may contain the visual information that users attune to in perceiving aural depth. Overall, our findings contribute to the literature on medium-field auditory depth perception research in VR, providing insights into the development of more immersive and effective VR experiences that support better perceptual outcomes.

The findings obtained in this study open the floor to a number of interesting follow-up questions. How long do the effects of calibration persist in users? Can calibrative feedback provided aurally also facilitate perceptual learning? Would aural depth perception be the same under a totally dark environment? In the future, we wish to obtain answers to such questions. Our immediate interests, however, lie in determining how aural depth perception occurs when the environment is totally dark. If having information about the extent of the room is sufficient to be precise in auditory depth perception, then a totally dark environment that does not afford this information should degrade perceptual accuracy. We wish to obtain insights into this by studying how users perceive the depth of an aural stimulus when not afforded any information about the room's extent.

Acknowledgments

This work was partly supported by the Taiwan National Science and Technology Council under Grant No. 109-2221-E-009-123-MY3, and the US National Science Foundation (CISE IIS HCC) under Grant No. 2007435.

REFERENCES

[1] B. M. Altenhoff, P. E. Napieralski, L. O. Long, J. W. Bertrand, C. C. Pagano, S. V. Babu, and T. A. Davis. Effects of calibration to visual and haptic feedback on near-field depth perception in an immersive virtual environment. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 71–78, 2012.

[2] B. M. Altenhoff, C. C. Pagano, I. Kil, and T. C. Burg. Learning to perceive haptic distance-to-break in the presence of friction. *Journal of Experimental Psychology: Human Perception and Performance*, 43(2):231, 2017.

[3] J. Andre and S. Rogers. Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & Psychophysics*, 68(3):353–361, 2006.

[4] A. Andreasen, M. Geronazzo, and N. C. Nilsson. Auditory feedback for navigation with echoes in virtual environments: Training procedure and orientation strategies. *IEEE Transactions on Visualization and Computer Graphics*, 25:1876–1886, 2019.

[5] A. Bhargava, R. Venkatakrishnan, R. Venkatakrishnan, K. Lucaites, H. Solini, A. C. Robb, C. C. Pagano, and S. V. Babu. Can i squeeze through? effects of self-avatars and calibration in a person-plus-virtual-object system on perceived lateral passability in vr. *IEEE Transactions on Visualization and Computer Graphics*, 2023.

[6] A. Bhargava, R. Venkatakrishnan, R. Venkatakrishnan, H. Solini, K. Lucaites, A. C. Robb, C. C. Pagano, and S. V. Babu. Did i hit the door? effects of self-avatars and calibration in a person-plus-virtual-object system on perceived frontal passability in vr. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4198–4210, 2021.

[7] G. P. Bingham and C. C. Pagano. The necessity of a perception–action approach to definite distance perception: Monocular distance perception to guide reaching. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1):145–168, 1998.

[8] J. Blascovich, J. Loomis, A. C. Beall, K. R. Swinth, C. L. Hoyt, and J. N. Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological inquiry*, 13(2):103–124, 2002.

[9] J. J. Blau and J. B. Wagman. *Introduction to ecological psychology: A lawful approach to perceiving, acting, and cognizing*. Taylor & Francis, 2022.

[10] J. Broderick, J. Duggan, and S. Redfern. The importance of spatial audio in modern games and virtual environments. In *2018 IEEE Games Entertainment Media Conference (GEM)*, pp. 1–9, 2018.

[11] E. R. Calcagno, E. L. Abregú, M. C. Eguía, and R. Vergara. The role of vision in auditory distance perception. *Perception*, 41(2):175–192, 2012.

[12] C. I. Cheng and G. H. Wakefield. Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space. In *Audio Engineering Society Convention 107*, 1999.

[13] S. H. Creem-Regehr and B. R. Kunz. Perception and action. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6):800–810, 2010.

[14] B. Cullen, K. Collins, A. Hogue, and B. Kapralos. Sound and stereoscopic 3D: Examining the effects of sound on depth perception in stereoscopic 3D. In *2016 7th International Conference on Information, Intelligence, Systems Applications (IISA)*, pp. 1–6. IEEE, 2016.

[15] B. Day, E. Ebrahimi, L. S. Hartman, C. C. Pagano, A. C. Robb, and S. V. Babu. Examining the effects of altered avatars on perception-action in virtual reality. *Journal of Experimental Psychology: Applied*, 25(1):1, 2019.

[16] B. Dong, A. Chen, Z. Gu, Y. Sun, X. Zhang, and X. Tian. Methods for measuring egocentric distance perception in visual modality. *Frontiers in Psychology*, 13, 2022.

[17] E. Ebrahimi. *Investigating embodied interaction in near-field perception-action re-calibration on performance in immersive virtual environments*. PhD thesis, Clemson University, 2017.

[18] E. Ebrahimi, B. Altenhoff, L. Hartman, J. A. Jones, S. V. Babu, C. C. Pagano, and T. A. Davis. Effects of visual and proprioceptive information in visuo-motor calibration during a closed-loop physical reach task in immersive virtual environments. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 103–110, 2014.

[19] E. Ebrahimi, B. M. Altenhoff, C. C. Pagano, and S. V. Babu. Carryover effects of calibration to visual and proprioceptive information on near field distance judgments in 3D user interaction. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 97–104. IEEE, 2015.

[20] E. Ebrahimi, L. S. Hartman, A. Robb, C. C. Pagano, and S. V. Babu. Investigating the effects of anthropomorphic fidelity of self-avatars on near field depth perception in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1–8. IEEE, 2018.

[21] F. El Jamiy and R. Marsh. Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing*, 13(5):707–712, 2019.

[22] P. E. Etchemendy, E. Abregú, E. R. Calcagno, M. C. Eguia, N. Vechiatti, F. Iasi, and R. O. Vergara. Auditory environmental context affects visual distance perception. *Scientific reports*, 7(1):7189, 2017.

[23] A. Farnell. *Designing sound*. MIT Press, 2010.

[24] I. T. Feldstein, F. M. Kölsch, and R. Konrad. Egocentric distance perception: A comparative study investigating differences between real and virtual environments. *Perception*, 49(9):940–967, 2020.

[25] D. Finnegan. *Compensating for distance compression in virtual audio-visual environments*. PhD thesis, University of Bath, 2017.

[26] D. J. Finnegan, E. O'Neill, and M. J. Proulx. Compensating for distance compression in audiovisual virtual environments using incongruence. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 200–212. ACM, 2016.

[27] D. J. Finnegan, E. O'Neill, and M. J. Proulx. An approach to reducing distance compression in audiovisual virtual environments. In *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pp. 1–6. IEEE, 2017.

[28] S. S. Fukusima, J. M. Loomis, and J. A. Da Silva. Visual perception of egocentric distance as assessed by triangulation. *Journal of experimental psychology: Human perception and performance*, 23(1):86, 1997.

[29] D. A. Gajewski, J. W. Philbeck, P. W. Wirtz, and D. Chichka. Angular declination and the dynamic perception of egocentric distance. *Journal of Experimental Psychology: Human Perception and Performance*, 40(1):361, 2014.

[30] D. A. Gajewski, C. P. Wallin, and J. W. Philbeck. Gaze behavior and the perception of egocentric distance. *Journal of Vision*, 14(1):20–20, 2014.

[31] J. J. Gibson. *The ecological approach to visual perception: classic edition*. Psychology Press, 2014.

[32] W. C. Gogel. Convergence as a cue to absolute distance. *The Journal of Psychology*, 52(2):287–301, 1961.

[33] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, eds., *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, 1988.

[34] B. Hervy, F. Laroche, J.-L. Kerouanton, A. Bernard, C. Courtin, L. D'haene, B. Guillet, and A. Waels. Augmented historical scale model for museums: From curation to multi-modal promotion. In *Proceedings of the 2014 Virtual Reality International Conference*, VRIC '14. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2617841.2617843

[35] Y.-H. Huang, R. Venkatakrishnan, R. Venkatakrishnan, S. V. Babu, and W.-C. Lin. Using audio reverberation to compensate distance compression in virtual reality. In *ACM Symposium on Applied Perception 2021*, pp. 1–10, 2021.

[36] G. Kailas and N. Tiwari. Design for immersive experience: Role of spatial audio in extended reality applications. In *Design for Tomorrow—Volume 2: Proceedings of ICoRD 2021*, pp. 853–863. Springer, 2021.

[37] J. W. Kelly, J. M. Loomis, and A. C. Beall. Judgments of exocentric direction in large-scale space. *Perception*, 33(4):443–454, 2004.

[38] B. R. Kunz, L. Wouters, D. Smith, W. B. Thompson, and S. H. Creem-Regehr. Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking. *Attention, Perception, & Psychophysics*, 71(6):1284–1293, 2009.

[39] W.-Y. Lin, Y.-C. Wang, D.-R. Wu, R. Venkatakrishnan, R. Venkatakrishnan, E. Ebrahimi, C. Pagano, S. V. Babu, and W.-C. Lin. Empirical

evaluation of calibration and long-term carryover effects of reverberation on egocentric auditory depth perception in vr. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 232–240. IEEE, 2022.

[40] M. A. Livingston, J. E. Swan, J. L. Gabbard, T. H. Hollerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown. Resolving multiple occluded layers in augmented reality. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pp. 56–65. IEEE, 2003.

[41] J. M. Loomis, J. A. Da Silva, N. Fujita, and S. S. Fukusima. Visual space perception and visually directed action. *Journal of experimental psychology: Human Perception and Performance*, 18(4):906, 1992.

[42] J. M. Loomis, R. L. Klatzky, J. W. Philbeck, and R. G. Golledge. Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics*, 60:966–980, 1998.

[43] J. M. Loomis, J. M. Knapp, et al. Visual perception of egocentric distance in real and virtual environments. *Virtual and Adaptive Environments*, 11:21–46, 2003.

[44] J. M. Loomis and J. W. Philbeck. Measuring spatial perception with spatial updating and action. In *Embodiment, Ego-space, and Action*, pp. 17–60. 2008.

[45] P. Maruhn, S. Schneider, and K. Bengler. Measuring egocentric distance perception in virtual reality: Influence of methodologies, locomotion and translation gains. *PloS one*, 14(10):e0224651, 2019.

[46] S. Masnadi, K. Pfeil, J.-V. T. Sera-Josef, and J. LaViola. Effects of field of view on egocentric distance perception in virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–10, 2022.

[47] F. O. Matu, M. Thøgersen, B. Galsgaard, M. M. Jensen, and M. Kraus. Stereoscopic augmented reality system for supervised training on minimal invasive surgery robots. In *Proceedings of the 2014 Virtual Reality International Conference*, pp. 1–4, 2014.

[48] J. W. McCandless, S. R. Ellis, and B. D. Adelstein. Localization of a time-delayed, monocular virtual object superimposed on a real environment. *Presence*, 9(1):15–24, 2000.

[49] D. H. Mershon and J. N. Bowers. Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8(3):311–322, 1979.

[50] B. J. Mohler, S. H. Creem-Regehr, and W. B. Thompson. The influence of feedback on egocentric distance judgments in real and virtual environments. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization*, pp. 9–14, 2006.

[51] P. E. Napieralski, B. M. Altenhoff, J. W. Bertrand, L. O. Long, S. V. Babu, C. C. Pagano, J. Kern, and T. A. Davis. Near-field distance perception in real and virtual environments using both verbal and action responses. *ACM Transactions on Applied Perception (TAP)*, 8(3):1–19, 2011.

[52] D. Narciso, M. Bessa, M. Melo, A. Coelho, and J. Vasconcelos-Raposo. Immersive 360° video user experience: impact of different variables in the sense of presence and cybersickness. *Universal Access in the Information Society*, 18:77–87, 2019.

[53] M. Paquier, N. Côté, F. Devillers, and V. Koehl. Interaction between auditory and visual perceptions on distance estimations in a virtual environment. *Applied Acoustics*, 105:186–199, 2016.

[54] J. Pätynen, V. Pulkki, and T. Lokki. Anechoic recording system for symphony orchestra. *Acta Acustica united with Acustica*, 94(6):856–865, 2008.

[55] J. W. Philbeck and J. M. Loomis. Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1):72, 1997.

[56] K. Ponto, M. Gleicher, R. G. Radwin, and H. J. Shin. Perceptual calibration for immersive display environments. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):691–700, 2013.

[57] B. Raveendranath, C. C. Pagano, M. Nasiri, A. C. Robb, and S. V. Babu. Effect of texture on the perception of axis of rotation of rotating panels. *Ecological Psychology*, pp. 1–30, 2022.

[58] M. Rébillat, X. Boutillon, É. Corteel, and B. F. Katz. Audio, visual, and audio-visual egocentric distance perception by moving subjects in virtual environments. *ACM Transactions on Applied Perception (TAP)*, 9(4):1–17, 2012.

[59] J. J. Rieser, H. L. Pick, D. H. Ashmead, and A. E. Garing. Calibration of human locomotion and models of perceptual-motor organization. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3):480, 1995.

[60] J. P. Rolland, W. Gibson, and D. Ariely. Towards quantifying depth and size perception in virtual environments. *Presence: Teleoperators & Virtual Environments*, 4(1):24–49, 1995.

[61] J. P. Rolland, C. Meyer, K. Arthur, and E. Rinalducci. Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. *Presence*, 11(6):610–625, 2002.

[62] J. L. Semmlow and D. Heerema. The role of accommodative convergence at the limits of fusional vergence. *Investigative ophthalmology & visual science*, 18(9):970–976, 1979.

[63] R. Sousa, E. Brenner, and J. Smeets. A new binocular cue for absolute distance: Disparity relative to the most distant structure. *Vision Research*, 50(18):1786–1792, 2010.

[64] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 13(3):429–442, 2007.

[65] J. A. Thomson. Is continuous visual monitoring necessary in visually guided locomotion? *Journal of Experimental Psychology: Human Perception and Performance*, 9(3):427, 1983.

[66] A. Turner, J. Berry, and N. Holliman. Can the perception of depth in stereoscopic images be influenced by 3D sound? *Proc.SPIE*, 7863:1–10, 2011.

[67] Valve. Steam audio manual, 2023.

[68] C. Valzolgher, M. Alzhaler, E. Gessa, M. Todeschini, P. Nieto, G. Verdelet, R. Salemme, V. Gaveau, M. Marx, E. Truy, et al. The impact of a visual spatial frame on real sound-source localization in virtual reality. *Current Research in Behavioral Sciences*, 1:100003, 2020.

[69] S. van Andel, M. H. Cole, and G.-J. Pepping. A systematic review on perceptual-motor calibration to changes in action capabilities. *Human movement science*, 51:59–71, 2017.

[70] R. Venkatakrishnan, R. Venkatakrishnan, B. Raveendranath, C. C. Pagano, A. C. Robb, W.-C. Lin, and S. V. Babu. How virtual hand representations affect the perceptions of dynamic affordances in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2258–2268, 2023.

[71] K. V. von Fieandt. Visual factors in space perception, 2023. `https://www.britannica.com/science/space-perception/Visual-factors-in-space-perception`.

[72] W. H. Warren. The dynamics of perception and action. *Psychological review*, 113(2):358, 2006.

[73] Widex. Widex hearing test, 2021.

[74] R. Withagen and C. F. Michaels. The role of feedback information for calibration and attunement in perceiving length by dynamic touch. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6):1379, 2005.

[75] B. G. Witmer and P. B. Kline. Judging perceived and traversed distance in virtual environments. *Presence*, 7(2):144–167, 1998.

[76] B. Wu, T. L. Ooi, and Z. J. He. Perceiving distance accurately by a directional process of integrating ground information. *Nature*, 428(6978):73–77, 2004.

[77] P. Zahorik. Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4):1832–1846, 2002.