2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3315-1647-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISMAR62088.2024.00057

Room Size Perception in Virtual Reality by Means of Sound and Vision: The Role of Perception-Action Calibration

Dai-Rong Wu* National Yang Ming Chiao Tung Univ.
Tyler Duffrin† Clemson University
Roshan Venkatakrishnan‡ University of Florida
Rohith Venkatakrishnan§ University of Florida
Sabarish V. Babu ¶ Clemson University
Christopher Pagano‖ Clemson University
Wen-Chieh Lin** National Yang Ming Chiao Tung Univ.

Abstract

Spatial perception in virtual reality (VR) has been a hot research topic for years. Most of the studies on this topic have focused on visual perception and distance perception. Fewer have examined auditory perception and room size perception, although these aspects are important for improving VR experiences. Recently, a number of studies have shown that perception can be calibrated to information that is relevant to the successful completion of everyday tasks in VR (such as distance estimation and spatial perception). Also, some recent studies have examined calibration of auditory perception as a way to compensate for the classic distance compression problem in VR. In this paper, we present a calibration method for both visual and auditory room size perception. We conducted experiments to investigate how people perceive the size of a virtual room and how the accuracy of their size perception can be calibrated by manipulating perceptible auditory and visual information in VR. The results show that people were more accurate in perceiving room size by means of vision than in audition, but that they could still use audition to perceive room size. The results also show that during calibration, auditory room size perception exhibits learning effects and its accuracy was greatly improved after calibration.

Index Terms: Auditory Room Size Perception, Calibration, VR

1 Introduction

The perceived size of a virtual space occupied by a user in VR is crucial for providing users better immersion and specific training applications, such as architectural design and visualization. Room size perception is thought to be related to distance perception. In VR research, distance compression is a common perceptual problem [44, 7, 30]. Similar to distance, underestimation also occurs in room size perception in VR [58, 35]. Although room size perception is thought to be related to distance perception, there are studies in the real world showing that it could be affected by numerous other factors [46], and thus may be a more complex perceptual phenomena to study. However, compared to distance perception, there are fewer research focused on room size perception in VR.

While many studies have focused on visual information, fewer studies have centered on auditory information. However, auditory information is still crucial for us to be aware of the surroundings. For example, there are studies showing that audio information can affect our visual distance perception [56, 8, 16]. Those studies indicate that auditory information is also important, yet there are still few studies focused on auditory modality of perception in VR. Recently, it has been shown that individuals can use auditory information in Immersive Virtual Environments (IVEs) to become calibrated to the distance of a sound source. *Calibration* is the process by which a person's action capabilities become properly scaled to environmental properties [9]. Moreover, calibration takes place as individuals are provided with corrective feedback about the accuracy of their perceptual judgments. After individuals make a judgment, they are given feedback about the accuracy of their judgment. As they make similar judgments in the future, they use that feedback to improve the accuracy of their perceptuomotor behaviors. Many studies have demonstrated that calibration can be used to improve accuracy of distance judgments in the real world and IVEs [10, 55]. For instance, Mohler et al. [38] used various forms of feedback to improve distance estimation of walking in VR. Relatedly, Ebrahimi et al. [15] found a similar result when investigating depth judgments via physical reaches. Focusing specifically on calibration to auditory information in VR, Lin et al. [34] demonstrated that reverberation time can be used to perceive distance in an IVE; they reported that participants' accuracy improved in future trials after they were given feedback about their performance on distance estimation. Lin et al. [33] used a similar methodology but investigated the effects of environmental visibility and an action-based measures to calibrate distance perception in VR. Again, their study showed that participants became calibrated to the distance of the sound source after receiving feedback about their performance.

The presented study investigates calibration using auditory information, but instead of focusing on egocentric distance, it focuses on *virtual room size perception*. We contend that room size perception should be studied by VR researchers for the following reasons: First, room size perception is important for presence, immersion and feelings about the space in IVEs [49, 21], and is therefore critical for VR applications such as architectural visualization, interior design, and virtual tourism. Second, realistic sounds that are heard in the context of VR rooms could contribute to feelings of presence and immersion [52], which apply to most any virtual scenario. Third, individuals who participate in training simulations that involve low visibility (such as fire safety with smoke-filled rooms) will need to accurately perceive room size, based mostly on sound. Fourth, training simulations for individuals without vision could make use of such a paradigm. Fifth, room size perception is theoretically interesting; it is related to distance perception, but is more complicated [46]. While egocentric distance has been well-studied in VR (and can be corrected with calibration), we endeavored to examine whether people could perform a similar, but more cognitively complex task and be corrected through calibration.

In the current study, we extend this calibration paradigm into the domain of virtual room size perception. Rather than asking participants to perceive distance to an auditory event in an IVE, we ask them to perceive the size of the IVE they are in by means of auditory information, and give them visual information to calibrate their

*e-mail:azusa871227.cs10@nycu.edu.tw
†e-mail:tduffri@clemson.edu
‡e-mail:rvenkat@g.clemson.edu
§e-mail:rohithv@g.clemson.edu
¶e-mail:sbabu@clemson.edu
‖e-mail:cpagano@clemson.edu
**e-mail:wclin@cs.nctu.edu.tw

979-8-3315-1647-5/24/$31.00 ©2024 IEEE
DOI 10.1109/ISMAR62088.2024.00057

performance. To do so, we borrow the room re-creation method introduced by Gagnon et al. [19]. In their study, participants stood in a virtual room. After scanning the room and indicating that they were ready to estimate its size, all visual information disappeared. If the response was a verbal report or spaciousness rating, participants were placed back in the virtual room and asked to make a report. But, when they used the room re-creation method, participants were placed in an "adjustment room" and asked to adjust the walls until they felt that the adjusted room was the same size as the trial room. In our study, we utilized this room re-creation method, asking participants to perceive room size based on auditory information or visual information. In the process, we aimed to investigate whether users' auditory room size perception can be calibrated in a manner similar to visual room size perception.

Our study intends to make the following contributions: (1) To the best of our knowledge, this work is the first to investigate the calibration of virtual room size perception; (2) The proposed calibration methodology effectively induces perceptual learning in VR users; (3) We demonstrated that the perception of virtual space can be calibrated by means of auditory information. This is important because IVEs are defined in part by their spatial constitution, and auditory information is a crucial factor for feelings of presence and fidelity in VR experiences.

2 Related Work

2.1 Visual Distance and Room Size Perception

Visual distance perception. Visual room size perception is influenced by the dimensions of a room, and is thus related to distance perception. There is an extensive volume of studies related to distance perception in both the real world and VR. Distance perception can be classified as exocentric and egocentric distance perception by different reference frames. Exocentric distance refers to distance between objects, and egocentric distance refers to distance between an observer and an object. Both exocentric [2] and egocentric [44] distance perception have been studied in the real world and VR. Others have looked specifically at the factors that affect distance perception in VR [57, 42]. For example, Vienne et al. studied whether screen distances and potent display-related factors (resolution, display orientation, luminance non-uniformity, and specular reflection) influenced distance perception in VR systems [57]. Ping et al. designed different kinds of visual cues and studied their influences on depth perception in both VR and AR [42]. Besides, regarding the accuracy of perceived distance in VR, distance compression or underestimation in VR is a well-studied problem. It is a phenomenon in which users in VR simulations estimate the distance between themselves and surrounding objects closer than the actual (simulated) location. Renner et al. offered a thorough literature survey on the distance compression problem in VR [44]. More recent studies revealed that perceived distances with newer HMDs is more accurate, but underestimation still exists [4, 29].

Visual room size perception. Compared to distance perception, room size perception is less studied. Most studies on room size perception have investigated the parameters that influence room size estimation. For example, Sadalla et al.'s experiments demonstrated that in the real world, rectangular rooms were perceived to be larger than square rooms, when all rooms in question had the same area and volume. They called this phenomenon "the rectangularity illusion" [46]. This phenomenon also implies that visual room size perception is related to distance perception, but that it is slightly more complicated. Saulton et al. conducted a similar study in VR and found that this phenomenon exists for virtual rooms as well [48]. Besides the shape of a room, Castell et al. investigated how the arrangement of furniture in a room influenced perceived spaciousness in true-to-scale model rooms and in VR [58]. Simpson et al. examined how different wallpaper influences room size perception in VR [50]. Stamps et al. conducted three experiments on how horizontal area, boundary height, elongation, and color affected perceived spaciousness[51]. Like distance perception, visual room size perception is also underestimated in VR [58, 35]

2.2 Auditory Distance and Room Size Perception

Auditory distance perception. Previous work has investigated influences of different audio cues on distance perception, for example, sound level, direct-to-reverberant (DRR) energy ratio, sound spectral shape, binaural cues, and dynamic cues [31]. Distance compression also exists in auditory distance perception. Zahorik et al. conducted a series of experiments examining humans' abilities to perceive sound source distance in a real-world room. They were interested in how listeners combine and weight multiple distance cues[63]. Their findings indicate that humans tend to underestimate sound source distance, and that the extent of underestimation was exponentially related to the actual source distance. In general, auditory distance perception is underestimated in the far field, and is more accurate in the near field–but generally most accurate at the distance that is approximately 1 m from the listener [17, 27, 63, 64].

Auditory room size perception. In studies on auditory room size perception, measures such as reverberation time, clarity index, and the DRR energy ratio have been used. For example, Hameed et al. compared accuracy of room size perception when reverberation time and DRR were used to measure room size perception. They found that reverberation time was more useful than DRR [22]. Yadav et al. convolved the audio stimuli with different oral-binaural room impulse responses and let participants rate the spaciousness of the room they perceived. They found that reverberation time was the strongest predictor for room size judgment[62]. In a study by Maempel et al., participants were asked to estimate distance to a sound source and room size in audio, visual, and audio-visual conditions with binaural room impulse responses (BRIR) and stereoscopic images from four rooms, and found that participants generally underestimated the size of the room[36].

2.3 Calibration of Spatial Perception

Calibration, or perceptual learning, aims to ameliorate the inaccuracy problem of perception, such as distance compression. Richardson et al. provided error correction feedback as perceptual information to participants, and found that they were able to calibrate with the information and had improved accuracy in a blind walking task [45]. Ebrahimi et al. examined calibration to near-field distance judgments and showed that participants were able to calibrate physical reach in the presence of visual and proprioceptive feedback [12]. Lin et al. used exaggerated reverberation time as audio stimuli and calibrated participants' visual distance perception, and showed that calibration effects lingered for six months after initial trials [34]. Although there are many studies on calibration of distance perception, calibration of room size perception - either visually or aurally - has rarely been addressed.

2.4 Measures of Perceived Room Size

Three measures of perceived room size are commonly used in studies on spatial perception: magnitude comparisons [24, 46], spaciousness ratings [62, 25], and verbal reports [32]. Magnitude comparisons and spaciousness ratings can provide a holistic estimation for all of a room's dimensions, but they cannot be easily converted into the exact size perceived. While verbal report provides the exact size, it could cause more variance and hence be less reliable [19]. Action-based measures (such as those used by Simpson et al. [50]) are useful, but impractical for the perception of large spaces. Gagnon et al. proposed a novel measure for room size perception in VR called *room re-creation* [19]. Their study indicated that this measure captures holistic perception of room size and is more accurate for size perception than verbal report.

3 Research Questions and Hypotheses

This study aimed to answer the following research questions: "How does the accuracy of the perception of room size vary under different measures and different types of sensory information (conditions)", and "How does the calibration work in different measures and conditions?". Based on these research questions, we developed the following hypotheses.

H1: Room re-creation as a measure of perceived room size produces more accurate perceptual judgements than verbal report.
H2: Room size perception will be more accurate in the visual condition than the auditory condition.
H3: Calibration will successfully improve the accuracy of room sizes perception both aurally and visually.

Most previous studies on spatial size perception instructed participants to verbally report visually perceived room size [32], which may have caused considerable variance [19]. In this study, we also asked participants to complete a room re-creation task [19] to measure perceived room size, and we expected that participants' judgments in the room re-creation task will be more accurate.

Previous studies investigated room size perception in terms of visual [50, 19] and auditory sensory channels [22, 62], looking at them in isolation. However, very few of them compared the differences of room size perception between these two modalities. As most people's visual sense is more dominant than their auditory sense in perceiving stimuli, it is expected that participants' room size perception would be more accurate in the visual condition.

Previous studies demonstrated that perceptual judgments in VR (focusing on sound or vision) can be improved through training or calibration [43, 11, 34, 1, 26]. However, to the best of our knowledge, research on calibration of room size perception–either visual or auditory–has rarely been done. Given the successful demonstrations in the aforementioned studies, we expect users' performance to improve after undergoing a calibration phase.

4 EXPERIMENT

4.1 System Description

We used an HTC Vive Pro HMD to conduct the experiments. ATH-M20x headphones with noise cancellation were used to render the audio stimuli in the audio condition. In the visual condition, the participants only wore the HMD with built-in headphones without playing any audio stimuli. The participants performed the required tasks with an HTC Vive controller. The experiment program was developed with Unity (Version 2019.4.4) and run on a PC with a 2.5GHz Intel i7 processor, 32GB RAM, an NVIDIA GeForce RTX 3070 graphics card, and Windows 10 operating system.

Audio Stimuli in the virtual scenes were generated by Steam Audio Engine [6]. An anechoic horn sound [41] was blended with the auralization setting and repeated in the virtual scene for perceiving the audio stimuli. Steam Audio simulates sound spatialization and propagation in real time. It features various auralization parameters, including customized HRTF fast convolution, and several setting options for reflections and occlusion regarding different kinds of raycasting, air absorption, and physics-based attenuation [1]. This work simulated auralization of different room sizes. The physical-attenuation and air-absorption indices were fixed so that the differences of the audio stimuli were due to the room size changes but not from physical attenuation and air absorption.

The materials of the walls, ceiling, and floor in Stream Audio were set as brick, wherein the scattering index was set as 0.05 and the transmission index 0.015 for all frequencies. The absorption index was set as 0.03, 0.04, and 0.07 for low, medium, and high frequencies, respectively. The central frequency for low, medium, and high frequency band was 400Hz, 2500Hz, and 15000Hz.

4.2 Virtual Scenes

Virtual Scene for Perceiving the Audio Stimuli (SA). Fig. 1(a) shows the virtual scene for participants to perceive audio stimuli. It is a dark room in which participants could only see a stereo speaker with a size of 0.5m × 0.5m × 0.5m and a yellow half-sphere with a radius of 5cm on the floor. The speaker was the sound source positioned in front of the participants with a distance of 1m and a height of 1.5m. The yellow sphere placed in the center of the room guided participants to position themselves in the center of the room. This setting ensures that sound reflections from each wall are almost the same for the participants, and the sound source roughly at the height of the participants' ears.

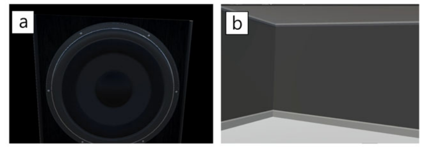

Figure 1: (a) Virtual scene SA: for participants to perceive the audio stimuli in a dark room with a stereo speaker. (b) Virtual scene SV: for participants to perceive the visual stimuli.

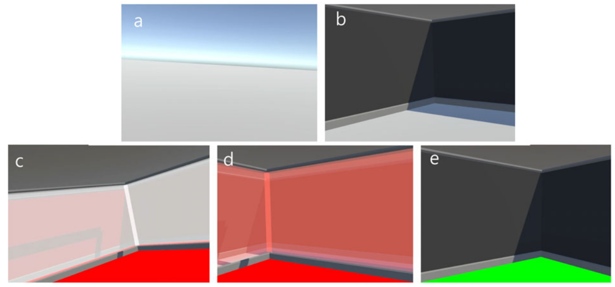

Figure 2: (a)(b) show the virtual scene SM, when a participant is doing verbal report and room re-creation, respectively. (c)-(e) show the virtual scene SC in the calibration phase. The red walls and floor indicate the actual (correct) room size. (c) and (d) show a participant's answer is underestimated and overestimated, respectively.

The room's height was fixed at 3 meters, and its width and length were the same. In the pretest phase and the post-test phase, the length of the room varied at five levels: 4m, 8m, 12m, 16m, and 20m. In the calibration phase, the length of the room varied at five levels: 6m, 10m, 14m, 18m, and 22m. This design avoids participants from simply memorizing the association of a stimulus and its corresponding room size, and the calibration effect can be evaluated without being biased by the memory effect of the participants.

Virtual Scene for Perceiving the Visual Stimuli (SV). Fig. 1(b) shows the virtual scene for the participants to perceive the room size in the visual condition. The participants were positioned in the center of an empty room. They could see the walls and ceiling without any audio stimuli. The settings of the room sizes in each phase were the same as those in SA since we would like to compare auditory and visual perception of room size.

Virtual Scenes for Measurement (SM). Initially, only the sky box and the floor were shown to the participants (Fig. 2(a)) as we don't want their verbal report affected by any other visual stimuli. Once the participants completed a verbal report, they would provide another measurement using the room re-creation method, which is started by showing walls and ceiling to participants, and they could move the walls with an HTC Vive controller (Fig. 2(b)). The walls would move by 0.5 meters each time the participant presses the button. The initial positions of walls were randomly set either 1m or 15m from participants with a uniform distribution so that they wouldn't cause any biases to the participants' perceived size measurements.

Virtual Scenes for Correction (SC). In the calibration phase, the actual room size would be shown to the participants after they confirmed their answer with the room re-creation measure. The actual (correct) room size was displayed by the red walls and floor. If a participant's answer was smaller than the actual size (i.e., underestimation), the walls re-created by the participant (i.e., representing perceived room size) would be shown in transparent (Fig. 2(c)). Reversely, if the participant's answer was larger than the actual room size (i.e., overestimation), the walls of the actual room would be transparent (Fig. 2(d)). This design allows participants to see both the perceived room and the actual room and to compare their difference. The calibration procedure repeats until the participant's answer is correct (the participant also move the walls in the same way as in SM), when the walls are shown in the original color (dark gray) and the floor will turn green (Fig. 2(e)), which informs the participant that the correct answer was hit.

4.3 Study Design

To study the perception of room size and its calibration with different types of sensory information and measures, we employed a 2 (Sensory Information Condition) × 2 (Reporting Method) × 3 (Phases) × 5 (Room Sizes) multi-factorial design, manipulating the sensory information as a between-subjects factor, the measurement method, phase, and room size as within-subjects factors. Two measures of perceived room size, verbal report and room re-creation, were tested to compare their reliability. Two types of sensory information were studied. Participants were randomly assigned to either of the two sensory conditions with roughly equal distributions of age, gender, and VR experience. In both conditions, they performed two room size judgment tasks in 3 sequential phases: pretest, calibration, and post-test. In the pre and post-test phases, users performed open-loop room size estimation and in the calibration phase, they performed closed-loop room size estimation. Similar to the methods used in[34, 33], the calibration phase allows users to observe the offset between the actual room size and their perceived room size, thereby allowing for correction of their estimation. In each phase, the trials featured 5 room sizes, each of which was repeated five times, thus adding up to a total of 25 trials. To avoid order effects, the order of these 25 trials was randomized. With each participant undergoing all three phases, this study involved each user making room size estimation for a total of 75 trials in each measurement type.

4.4 Participants

An apriori power analysis using G*Power was conducted to determine an appropriate sample size for our study. Based on an effect size of 0.20, an alpha error probability of 0.05, power of 0.95, 2 between-subjects groups (audio vs vision), 2 within-subjects measurement types (room recreation vs verbal report) × 5 room sizes × 5 trials per room size × 3 sessions (pretest, calibration and post-test) = 150 total measurements per participant, correlation of 0.5 among repeated measures and nonsphericity correction of 1, resulted in a total sample size of 16 participants (8 in each sensory information condition). We ultimately recruited a total of 42 participants, which is more than the number required by the G*Power analysis because this helps increase the diversity of the subject pool in terms of aspects like gender, age, game experiences, and educational background, thereby allowing for a more representative sample. Participants' ages ranged from 21 to 40 years old ($M = 23.3$, $SD = 2.88$). The participants are recruited through our university system. Most of the participants did not have virtual reality experience or had experienced virtual reality for less than five hours. All participants had normal or corrected-to-normal (20/20) vision and normal hearing acuity in both ears. In the audio condition, 20 participants (10 male, 10 female) were recruited and two of them were excluded as outliers as explained in Section 5 in detail. In the visual condition, we had 22 participants (12 male, 10 female). One of them dropped out due to cybersickness. In total, we had 18 and 21 participants in the audio and visual conditions for data analysis, respectively.

4.5 Tasks

Open-loop room size estimation (Pretest and Post-test Phases): Fig. 3 shows the process of the open-loop room size estimation. In the beginning of an experimental trial, a participant would be placed in the center of virtual scene SA (or SV) according to the condition s/he was assigned. In the audio condition, the participant listened to the audio stimuli produced from the speaker (Fig. 1(a)). In the visual condition, the participant observed the room size with eyes (Fig. 1(b)). After perceiving the stimuli, the participant pulled the trigger on the controller, and was transported to virtual scene SM (Fig. 2(a)). Next, the participant needed to verbally report the perceived room size. After the verbal report, the participant provided another measurement of perceived size using the room re-creation method (Section 4.7).

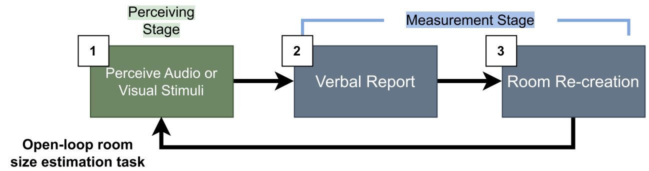

Figure 3: Flowchart of the open-loop room size estimation task.

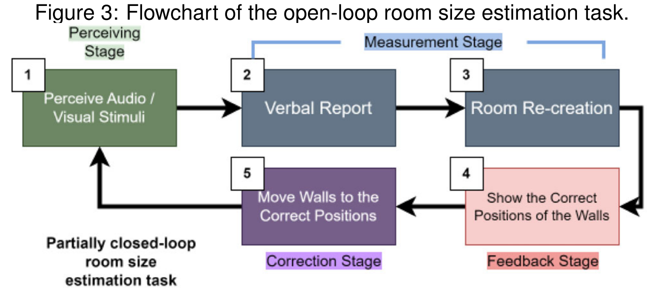

Figure 4: Partially closed-loop room size estimation task.

Closed-loop room size estimation (Calibration Phase): Fig. 4 shows the process of closed-loop room size estimation. A participant was positioned in the center of the virtual scene SA (or SV) and perceived the audio (or visual) stimuli in the beginning of a trial according to the condition s/he was assigned. After doing the verbal report and the room re-creation task, the actual/correct room size was shown. The participant was then asked to move the walls of the perceived room to the correct positions (Fig. 2(c)(d)). We took this procedure including perceiving feedback and correction as calibration.

4.6 Procedure

Upon arrival, participants were asked to read and sign a consent form (informed consent). After consenting to participate, participants completed a demographic survey that included information about their backgrounds and experiences with VR and video games. Following this, participants' interpupillary distances (IPD), eye heights, arm lengths, and stereo acuity were measured. They were then randomly assigned to one of the two sensory conditions. The experimenter then detailed the task they would be performing. For participants in the audio condition, their ear anthropometric parameters were also measured for constructing their personal HRTF. They then proceeded to take the Widex hearing test [60].

After receiving instructions, participants performed a few practice trials to familiarize themselves with the mechanics of the task. In practice trials, participants looked at the computer screen and practiced the task without wearing the HMD. After practice trials, participants put on the HMD and went through pretest, calibration, and post-test phases in order. In the pretest and post-test phases, users performed open-loop room size estimation, while performing closed-loop room size estimation in the calibration phase. They were allowed to take a short break between the pretest and calibration phase. During the experiment, participants stood in a fixed position, but they could still turn their upper body or head to perceive the stimuli. At the end of the experiment, participants filled out the NASA TLX questionnaire. Finally, they engaged in a short semi-structured interview, after which they were debriefed and compensated for their time. On average, it took about 90 minutes to complete the whole procedure.

4.7 Measures

Perceive room size (PS) was measured using verbal reports and room re-creation methods: (1) **Verbal report**: Participants had to estimate the distance between the walls in front of and behind them. They reported this distance in terms of the number of arm lengths between the two walls. Their answers were converted to meters and recorded. (2) **Room re-creation**: Each participant had to stand at the center of the room in the measurement stage and estimate the size of a room by moving the walls of a virtual room. That is, they re-created a room whose size is what they perceived by using the HTC controller. We recorded the final position of the walls from participants' answer, and computed the side length of the re-created room. This method was proposed by Gagnon et al. [19] and a similar measure was also used in the study by Frank et al. [18]. We adopted this method of measurement because it is considered more accurate [5]. PS and **Actual Room Size (AS)** were described in terms of room length rather than volume because the room was square shaped with a fixed height. This was inspired by Gagnon et al.'s work [19] wherein they used room area to describe a room's size, only changing the length and width of the virtual room in their experiment. Notwithstanding room sizes being described in terms of length, auditory information heard by the users was modeled as a function of the room's geometry, affecting how the sound signals reflected of the walls of the room. This lends itself to the assertion that room sizes were indeed the aspect being investigated in this work rather than being simply based on distance.

Judgment time (JT) was operationalized as the time before which participants made a judgment of room size. It was computed as the time taken from the onset of the stimuli until the time at which the judgement was made.

NASA Task Load Index (NASA-TLX) was collected after all trials were completed. It measures participants' self-ratings on 6 factors (Mental, Physical Demand, Temporal Demand, Performance, Effort, Frustration) contributing to their perceived workloads [23]. Ratings are multiples of 10 on a scale from 0 to 100.

5 Results

We conducted a multiple linear regression to evaluate how participants' room size perception was affected by different sensory information conditions, phases, reporting methods, actual room sizes, and other interaction terms. Furthermore, we conducted repeated measure ANOVA tests to compare the signed error of room size perception under different conditions, phases, methods, and trials. For judgment time, another repeated measure ANOVA was performed to analyze the difference between phases and conditions. To understand the workload of the participants in the experiments, the paired sample t-test was performed on each experimental factor to check if there was any significant difference between different conditions.

5.1 Modeling Performance of Room Size Perception

Before the multiple regression analysis, an analysis of standardized residuals was conducted on the collected data to identify outliers. The data that were beyond ± 3.0 of the standardized residuals were dropped. Tests to ensure if the data met the collinearity assumptions indicated that multicollinearity was not a concern. The histogram of standardized residuals indicated that the errors of the data were normally distributed, as did the P-P plot of standardized residuals, which showed that the data was close to a linear regression profile. The standardized residuals scatter plot also indicated that the data conformed to the assumptions of variance and linearity, and the data also met the assumptions of non-zero variance.

A multiple regression model was performed to investigate whether the independent variables (AS, $Condition$, $Phase$, $ReportMethod$) predicted PS. A significant equation was found $F(4,1164) = 608.409$, $p < 0.001$, with $R^2 = 0.677$: $PS = 1.051m + 0.587 \times AS + 0.550 \times Condition + 1.094 \times Phase - 0.665 \times ReportingMethod$. For Condition, the audio condition was coded as 13, and visual condition was 17. Phase 1 was pretest, phase 2 was calibration, and phase 3 was post-test. Reporting Methods were coded as 11 for the room re-creation measure, and 19 for the verbal report measure. According to the model, PS increased by 0.587 meters for every meter of increase in AS. PS increased by 0.550 meters for a change in condition. PS increased 1.094 meters for changing through a phase to the next one (i.e., pretest to calibration, calibration to post-test). PS decreased by 0.665 meters for a difference in the Reporting Method. All of the four factors are significant variables to predict PS (AS, $p < 0.001$; Condition, $p < 0.001$; Phases, $p < 0.001$; Reporting Method, $p < 0.001$).

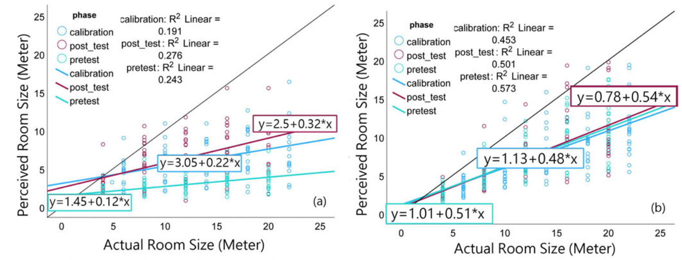

Figure 5: Linear regression results from the verbal report measure in the (a) audio and (b) visual conditions.

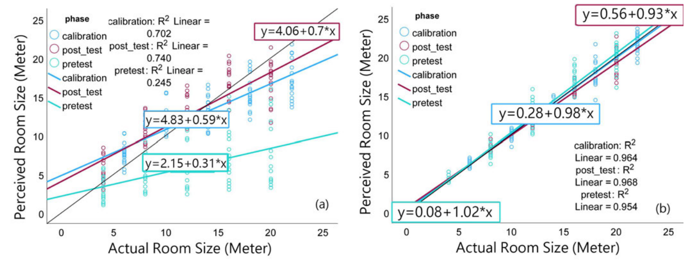

Figure 6: (a) Linear regression results from the room re-creation measure in the (a) audio and (b) visual conditions.

To evaluate the significant interaction effects, the independent variable AS was mean centered to eliminate any multicollinearity effects, and all the two-way, three-way, and four-way interaction terms were added to the model and a hierarchical multiple regression was performed. The regression model with the interaction variables was found to be significant, F(15,1164) = 321.061, $p < 0.001$, with an $R^2 = 0.807$, and all of the variables including interaction terms were significant to the model. For the detailed information, please refer to the supplementary materials.

The simple regression profiles for the participants' PS in each phase for the audio and visual conditions measured by different measures are shown in Fig. 5 and Fig. 6. The regression equations and enlarged plots are provided in the supplementary materials.

5.2 Signed Error of the Room Size Perception Between Conditions, Phases, and Reporting Methods

We investigated the calibration effect by comparing the signed errors ($SignedError = PS - AS$) in the pretest and post-test phase and the learning effect during the calibration phase. A positive signed error indicates that the participant overestimated the room size, while a negative signed error means an underestimation.

5.2.1 Signed Error from Trials in Pretest and Post-test

We performed a 2 (Conditions) $\times$ 2 (ReportMethods) $\times$ 2 (Phases: pretest, post-test) mixed model ANOVA to analysis - after verifying all the assumptions were met, namely normality, sphericity, and equality of variance. We only considered the pretest and post-test phases to examine the effect of learning before and after calibration. For the detailed results please see Table 1.

To further explore the learning effect in the overall 3 way interaction, we performed appropriate post-hoc pairwise comparisons on $Phase \times ReportingMethod \times Condition$ interaction effect using the Bonferroni method. In one set of 3 way comparisons on

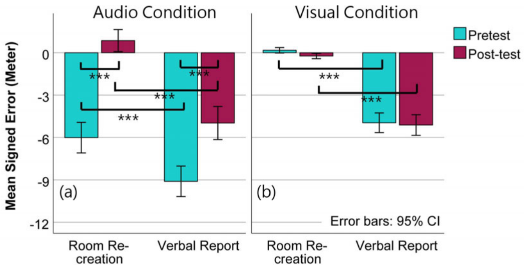

Figure 7: Mean signed error in the pretest and the post-test phase from the audio condition and the visual condition measured by the room re-creation measure and the verbal report measure. The significance within each condition was marked in the graph.

Main Effects (P: Phase, R: Reporting Method, C: condition)	
Within-subjects Variables	
P	$F(1,193) = 193.020$, $p < 0.001$, $\eta^2 = 0.500$
R	$F(1,193) = 130.983$, $p < 0.001$, $\eta^2 = 0.404$
Between-subjects Variables	
C	$F(1,193) = 42.541$, $p < 0.001$, $\eta^2 = 0.181$
Interaction Effects	
P×C	$F(1,193) = 161.137$, $p < 0.001$, $\eta^2 = 0.455$
P×R	$F(1,193) = 15.050$, $p < 0.001$, $\eta^2 = 0.072$
P×C×R	$F(1,193) = 21.638$, $p < 0.001$, $\eta^2 = 0.101$

Table 1: Main and interaction effects of RM-ANOVA in Section 5.2.1.

the signed error, in the audio condition, in the room re-creation method, the pretest phase ($M = -6.009, SD = 0.378$) was significantly lower than the post-test phase ($M = 0.856, SD = 0.278$), $p < 0.001$, and the signed error in the audio condition in the verbal report method was also significantly lower in the pretest phase ($M = -9.102, SD = 0.462$) than in the post-test phase ($M = -4.977, SD = 0.495$), $p < 0.001$.

In a different set of 3 way comparisons on signed error, post-hoc pairwise comparisons using Bonferroni method revealed the following. In the audio condition, in the pretest phase, the room re-creation method ($M = -6.009, SD = 0.378$) was significantly higher than the verbal report method ($M = -9.102, SD = 0.462$) $p < 0.001$. Also, in the audio condition, in the post-test phase, the room re-creation method ($M = 0.856, SD = 0.278$) was also significantly higher than the verbal report method ($M = -4.977, SD = 0.495$) $p < 0.001$. Similarly, in the visual condition, in the pretest phase, the room re-creation method ($M = 0.173, SD = 0.350$) was significantly higher than the verbal report method ($M = -4.958, SD = 0.428$) $p < 0.001$. Also, in the visual condition, in the post-test phase, the room re-creation method ($M = -0.235, SD = 0.257$) was significantly higher than the verbal report method ($M = -5.118, SD = 0.459$) $p < 0.001$.

In yet another set of 3 way comparison on signed error, post-hoc pairwise comparisons using Bonferroni method revealed the following. In the pretest phase, in the room re-creation method, the audio condition ($M = -6.009, SD = 0.378$) was significantly lower than the visual condition ($M = 0.173, SD = 0.350$) $p < 0.001$. In the pretest phase, in the verbal report method, the audio condition ($M = -9.102, SD = 0.462$ was significantly lower than the visual condition ($M = -4.958, SD = 0.428$) $p < 0.001$. Similarly, in the post-test phase, in the room re-creation method, the audio condition ($M = 0.856, SD = 0.278$) was significantly higher than the visual condition ($M = -0.235, SD = 0.257$) $p = 0.004$ (see Fig. 7).

5.2.2 Signed Error from Trials in Calibration Phase

To examine the learning effect in trials of the calibration phase, we conducted a 2 (Conditions) × 2 (ReportMethods) × 25 (Trials) mixed model ANOVA analysis on data from calibration phase only - after verifying all the assumptions were met, namely normality, sphericity, and equality of variance. Table 2 shows the results.

To further explore the learning effect of calibration phase between different conditions, we performed appropriate post-hoc pairwise comparisons on $Trial \times Condition$. We did not find any significant effects in the visual condition; however, in the audio condition, post-hoc pairwise comparisions revealed that $Trial$ 1 ($M = -8.481, SD = 0.956$) is significantly lower than in several subsequent trials, including $Trials$ $3,7,10,11,14,16,18,19,21,22,25$. $Trial$ 17 ($M = -5.271, SD = 0.687$) is significantly lower than $Trial$ 22 ($M = -1.700, SD = 0.818$), $p = 0.045$.

Main Effects (T: Trial, R: Reporting Method, C: condition)	
Within-subjects Variables	
T	*Greenhouse-Geisser adjusted* $F(11.917, 440.936) = 2.408$, $p = 0.005$, $\eta^2 = 0.061$
R	$F(1,37) = 173.528$, $p < 0.001$, $\eta^2 = 0.824$
Interaction Effects	
T×C	*Greenhouse-Geisser adjusted* $F(11.917, 440.936) = 1.966$, $p = 0.026$, $\eta^2 = 0.050$

Table 2: Main and interaction effects of RM-ANOVA in Section 5.2.2.

Main Effects (P: Phase, C: condition)	
P	*Greenhouse-Geisser adjusted* $F(1.509, 55.823) = 45.017$, $p < 0.001$, $\eta^2 = 0.549$
Interaction Effects	
P×C	*Greenhouse-Geisser adjusted* $F(1.509, 55.823) = 3.629$, $p = 0.045$, $\eta^2 = 0.089$

Table 3: Main and interaction effects of RM-ANOVA in Section 5.3

5.3 Judgement Time Between Conditions and Phases

To assess the average judgment time under different conditions and phases, we conducted a two-way mixed model ANOVA analysis with 2 $Condition$ × 3 $Phase$ factorial design on the participants' average judgment time, after verifying all the assumptions were met - namely normality, sphericity, and equality of variance. For the detailed results, please see Table 3.

To examine the $Phase$ by $Condition$ interaction further, appropriate post-hoc pairwise comparisons were performed. Post-hoc pairwise comparisons using the Bonferroni method revealed that, within the audio condition, the judgment time in the pretest phase ($M = 16.858s, SD = 2.618$) was significantly longer than the judgment time in the post-test phase ($M = 10.991s, SD = 1.753$), $p = 0.001$, and the judgment time in the calibration phase ($M = 13.762s, SD = 2.376$) was significantly longer than the judgment time in the post-test phase, $p = 0.007$. In the visual condition, all phases were significantly different from each other. Post-hoc pairwise comparisons using Bonferroni method revealed that the pretest phase ($M = 23.949s, SD = 2.424$) and was significantly longer than calibration phase the second longest ($M = 18.223s, SD = 2.2$), $p < 0.001$, and the post-test phase ($M = 13.442s, SD = 1.623$), $p < 0.001$. Post-hoc pairwise comparisons also revealed that the calibration phase was significantly longer than the post-test phase, $p < 0.001$. Post-hoc pairwise comparisons using Tukey's HSD method did not reveal any significant differences between the visual and audio conditions in the three phases (see Fig. 8).

5.4 Workloads Between Conditions

To compare the difference of perceived workload, we conducted a Mann-Whitney U test to evaluate each factor in NASA TLX questionnaire between the audio and the visual condition. The results indicated that Temporal Demand was significantly lower in the audio condition ($Mdn = 0$, $SD = 0.628$) than in the visual condition ($Mdn = 0$, $SD = 4.445$), $U = 135.00$, $Z = -2.156$, $p = 0.031$. Frustration was significantly higher in the audio condition ($Mdn = 18.333$, $SD = 6.375$) than in the visual condition ($Mdn = 6.667$, $SD = 8.878$), $U = 83.500$, $Z = -2.989$, $p = 0.003$. No other significant differences were found between the conditions.

6 Discussion

6.1 Verbal Report vs. Room Re-creation

The regression analyses revealed the existence of a significant relationship between the perceived size of the room and its actual size

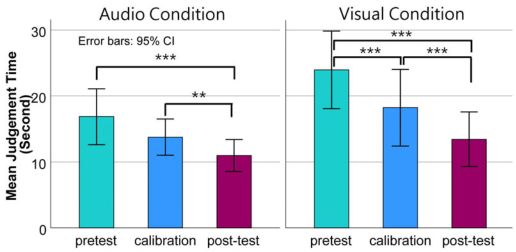

Figure 8: Mean judgment time in each phase and condition. The significance within each condition was marked on the figure.

with both measures. The model shows how users' perceptions of the room's size increases linearly with the stimuli presented (i.e., the room's actual size) (Fig. 5 and Fig. 6). A large proportion of variance in the perceived size is explained by the regression models, suggesting that participants are also highly consistent in their judgements. Further analyses on the signed error scores revealed that the room-recreation measure produced more accurate perceptual judgements than the verbal report. This can be inferred from Fig. 7 which shows the degree of overestimation or underestimation of the room's size across phases with each measure. Given that signed error scores are closer to 0 in the room-recreation measure than in verbal report, the former method produces more accurate perceptual judgements of room size than the latter. These results support **H1** which predicted that participants' accuracy will be greater in the room re-creation measure than in verbal reports. Our results directly align with those obtained by Gagnon et al. [19], showing evidence for the increased reliability of room re-creation as a measure of perceived room size in IVEs. On further inspecting Fig. 5 and Fig. 6 which depict the effects of actual distance on perceived distance moderated by the phase, we drew comparisons between the two measures in terms of the degrees of calibration induced. Towards this end, we interpreted how the regression lines of the different phases compared with the veridical judgement lines in black in each of these figures. These veridical lines represent the prediction where participants' perceived room size is exactly equal to actual room's size featuring a slope of 1 and intercept of 0. A regression line that is closer to the veridical judgment line implies that perceived room size is more accurate. While the post-test lines are closer to the veridical lines than the pre-test lines in general, suggesting improvement in perceptual judgements through calibtation, the degree of change is more pronounced in the room re-creation method than the verbal report method. This is indicative of the former method inducing a greater degree of calibration than the latter, further serving to support hypotheses **H1**, in that the degree of improvement in perceptual accuracy is more when using room-recreation as a measure than verbal report.

This pattern of results is consistent with several work that demonstrate that verbal reports are less accurate than action-based measures such as physical reaching, either in the real world [40] or VR [39]. While the room re-creation method does not require as much bodily movement and motor-control-based responses like blind-walking [28], physical reaching [14, 13], or beanbag throwing [47], it still involves manipulation of reported room size through adjustments made on a controller, affording users with direct visual information of the size of the room they are reporting. In contrast, verbal reports do not provide any visual information about the size of the room being reported as a perceptual judgement, and involve more of a cognitive stream rather than a motor control-based judgement [20, 3, 37, 40]. This naturally makes the task of perceiving size more prone to inaccuracy when performed verbally. Our findings highlight the need for researchers assessing the perception of an environment's size to leverage measures that are more reliable like the room-recreation measure used in this study, and Gagnon et al. [19], instead of traditionally used verbal reports.

6.2 Visual versus Audio Room Size Perception

Results obtained from the mixed modal ANOVA analysis conducted on pre-test and post-test signed error values inform us about over/underestimation of room size perception. With respect to this analysis, positive values indicate an overestimation of room size and negative values indicate an underestimation of room size. Looking at Fig. 7, we can see that for both audio conditions (room re-creation and verbal report), in the pretest, participants underestimated the room size. Interestingly, in the post-test, participants slightly overestimated room size when they used the room re-creation method, but continued to underestimate room size with the verbal report method. This suggests that participants' accuracy improved from pre-test to post-test when they perceived room size based on auditory information, when using the room re-creation method, which also happened when using the verbal report, but to a lesser extent. This finding becomes more meaningful when we compare it with the visual conditions of the task. For both visual room re-creation and verbal report, there was no difference between pre-test and post-test accuracy. That is, with vision, participants' judgments of room size did not become more accurate over the course of the study (Fig. 7).

Taken together, these results indicate that auditory information can be used to calibrate room size perception judgments in IVEs; when people are exposed to auditory information in VR, asked to make a judgment about room size, and are given feedback about their judgments, they can use that auditory information to refine their future judgments of room size. Accuracy of perception can be improved in this way. At the same time, in the visual condition, accuracy did not improve from pre-test to post-test. This is likely because the task was simply easier for participants in the visual condition; they could plainly see their estimations of room size in the pre-test and post-test. In the audio condition, however, participants had no frame of reference in the pre-test. But once we provided them with visual feedback, they were able to attune to the relevant information in their surroundings, and ultimately increase the accuracy of their future estimations.

These findings support **H2**, which is that room size perception will be more accurate in the visual condition than the auditory condition, and that underestimation of room size would exist in both the visual and auditory conditions, with the magnitude of underestimation being larger in the latter. Moreover, these findings are significant as they contribute to the large literature on calibration or perceptual learning [59, 61]. They also contribute to the discussion on how to mitigate the negative effects of distance compression in VR by means of auditory information. A number of researchers have used audio-based techniques to reduce distance compression, such as manipulation of timing and type of sound, reverberation level of sound, and various combinations of sound and vision [56, 8, 16, 34]. Here, we extend the problem of distance compression to a closely related issue, namely room size perception, showing that aurally perceived room sizes can be calibrated. Since the successful navigation of virtual space is central to any kind of VR application, our findings can be used in a number of domains. A VR training simulation in which a firefighter needs to ascertain how large a burning room is, having limited visual information is one such example of where room size perception is performed with auditory information. In such a situation, providing feedback can calibrate perceptions of room size, improving the accuracy of the same.

6.3 Calibration Effects

Our hypothesis **H3** was that calibration would improve accuracy of visual and audio room size perception. The regression analyses revealed calibration effects over the course of the experiment. Again, as shown in Fig. 5 and Fig. 6, the regression lines of the pre-test phase have a shallower slope and are less close to the veridical line than the lines of the post-test phase. These results are indicative of the success in employing a calibration phase toward improving participants' perception of room size, thus supporting **H1** and **H3**.

Here, we can also reference Section 5.2.2 and Fig. 8. Section 5.2.2 represents the mean signed error of each trial in the calibration phase of the audio condition of our study. In this context, mean signed error represents over/underestimation in the same way it did with respect to Fig. 5 and Fig. 6; values above 0 indicate overestimation and values below 0 indicate underestimation. Section 5.2.2 shows that, in the calibration phase of the audio condition, participants' underestimations were greater when they made verbal reports than when they used the room re-creation method. The same is true for Section 5.2.2. Even when they made their estimations based on visual information, in the calibration phase of the study, participants underestimated room size more when they performed verbal reports than when they used the room re-creation method. Succinctly, this pattern of results suggests that calibration was effective at inducing the perceptual learning of room size perception for both sound and vision, more so when participants completed the room re-creation task than the verbal report. Additionally, Fig. 8 shows us that participants took a longer time to make estimations of room size when they were exposed to visual information than when they were exposed to auditory information. This could be because participants were trying to be more careful or precise when they were given visual information than when they were given auditory information to estimate room size. This figure also demonstrates that the calibration effect took place from pre-test to post-test in both the visual and audio condition. Specifically, participants' judgment time decreased from pre-test to calibration to post-test in both conditions.

The calibration effects we observed here fall in line with a number of previous studies on calibration in VR [9, 10, 15, 12, 34, 33]. It has been demonstrated in these studies that participants' perceptual accuracy can be improved over time as they are given corrective feedback about their performance on a given task. We show the same thing here; that participants' accuracy can improve over the course of their completion of a perceptual task after they are given corrective feedback. But in this context, we demonstrate that individuals can become calibrated to virtual room size perception by means of auditory information. While there have been a number of studies on auditory distance perception and auditory room size perception in both the real world and VR, to the best of our knowledge, this is the first time that calibration of auditory room size perception has been investigated, and compared and contrasted to calibration of visual room size perception. As mentioned earlier, there have been a number of different strategies employed by investigators to mitigate distance compression in VR. Researchers have focused on sound level, direct-to-reverberant energy, sound spectral change, binaural cues, and dynamic cues in attempts to overcome distance compression by means of auditory information [56, 8, 16]. Here, we demonstrate that calibration can be used as a novel means of training people to improve their accuracy in judgments of room size based on auditory information. Further, we showed that audio perception could be calibrated by visual stimuli, which indicates the possibility of cross-modality calibration.

6.4 Judgment time and Workload

Overall, participants took longer to make judgments in the visual condition of the study than the audio condition. This could be because participants, when estimating room size based on vision, had more information to make a final judgment than when they estimated room size based on sound. As mentioned in the introduction, visual information is (theoretically) more constant or ubiquitous than auditory information, in the real world and IVEs. This is what we mean by "more information" with vision than sound. Perhaps, because they could be more fine-grained with their estimation of room size based on vision, participants took longer than they did with auditory information. Also, results from judgment time across both modalities indicated that the amount of time participants took to make estimates decreased (Fig. 8) across the three phases of the study in both audio and visual conditions. This implies that participants became more familiar with the task and thus did not need as much time to make estimates. While amount of time required to complete the task decreased, in all conditions of the study, accuracy increased (however slightly).

In NASA TLX results, participants reported feeling more time pressure in the visual condition than in the audio condition. These results are corroborated by the results of judgment time. Also, participants exerted more effort in completing the audio condition than the visual one. This could reflect the unfamiliar nature of the task.

6.5 Limitations

In this study, we restricted the shape of the room to a square and without furnishing. As auditory perception of room size could be affected by the shape of a room [1, 26], this could be an avenue for future research. We intentionally designed the room to be very plain to minimize any features that could distract the participants or confound the results. Another limitation is the distribution of the participants. They were recruited from our university and had similar VR experiences. We'd like to recruit participants with more diverse background to increase the generalizability of the findings. Also, we adopted repeated measures ANOVA to perform multi-factorial designs, which is usually used for two-factorial designs. However, to the best of our knowledge, it is still the best way to analyze our experiment results, and there are also previous studies using repeated measures ANOVA for multi-factorial designs [53, 54]. Finally, with the development of VR technologies, the results might also be different since studies have shown that with different HMDs, there would be different ratios of spatial underestimation [4]. It is an interesting research direction to calibrate room size perception across different VR devices.

7 Conclusion

We investigated the difference between visual and audio room size perception in VR with two measures. We found that room size perception was more accurate and consistent in visual condition than in audio condition. We also found that room re-creation [19] was a more reliable measure of perceived room size for both visual and audio conditions than verbal report. This is significant since there have been a number of studies on distance and room size perceptions in VR, but until now, none of them have used the room recreation method for the purpose of assessing calibration by means of auditory information in VR. Finally, we found that individuals can become calibrated for perception of virtual room size by auditory information. That is, auditory information can be used to train people to perceive, with some accuracy, how large or small a virtual room is. We demonstrated that people can learn to perceive the size of a virtual room with greater accuracy after receiving feedback on their performance in the form of visual information. Our findings thus shed light on the nature of human perception as it takes place in IVEs. This study opens the floor to a number of interesting follow-up experiments, what if calibration was facilitated aurally? Would the same trends hold? What if there were objects, people, and other artifacts or entities in the room? Would this affect how sizes of virtual environments are perceived? Specifically, our immediate interests lie in empirically evaluating the effects of calibration realized through aural feedback for the perception of room size based on aurally specified information. Given the growing advancements in spatial sound related devices and techniques, it behooves designers to study this phenomenon in greater depth in an effort to make this technology more accessible to a wider range of users.

Acknowledgments

This work was supported in part by Taiwan NSTC under grant no. 109-2221-E-009-123-MY3 and The US National Science Foundation (CISE: IIS: HCC) under grant no. 2007435. Finally, we would like to thank all the participants.

References

[1] A. Andreasen, M. Geronazzo, N. C. Nilsson, J. Zovnercuka, K. Konovalov, and S. Serafin. Auditory feedback for navigation with echoes in virtual environments: training procedure and orientation strategies. *IEEE transactions on visualization and computer graphics*, 25(5):1876–1886, 2019. 3, 8

[2] J. A. Aznar-Casanova, E. H. Matsushima, N. P. Ribeiro-Filho, and J. A. Da Silva. One-dimensional and multi-dimensional studies of the exocentric distance estimates in frontoparallel plane, virtual space, and outdoor open field. *The Spanish Journal of Psychology*, 9(2):273–284, 2006. 2

[3] 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, 1998. 7

[4] L. E. Buck, M. K. Young, and B. Bodenheimer. A comparison of distance estimation in hmd-based virtual environments with different hmd-based conditions. *ACM Transactions on Applied Perception (TAP)*, 15(3):1–15, 2018. 2, 8

[5] B. Burnett, A. Neidhardt, Z. Cvetković, H. Hacıhabiboğlu, and E. De Sena. User expectation of room acoustic parameters in virtual reality environments. In *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–10. IEEE, 2023. 5

[6] V. Corporation. Steam audio, 2018. 3

[7] S. H. Creem-Regehr, J. K. Stefanucci, and B. Bodenheimer. Perceiving distance in virtual reality: theoretical insights from contemporary technologies. *Philosophical Transactions of the Royal Society B*, 378(1869):20210456, 2023. 1

[8] 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. 1, 7, 8

[9] B. Day, E. Ebrahimi, L. S. Hartman, C. C. Pagano, and S. V. Babu. Calibration to tool use during visually-guided reaching. *Acta psychologica*, 181:27–39, 2017. 1, 8

[10] 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. 1, 8

[11] A. G. de Siqueira, R. Venkatakrishnan, R. Venkatakrishnan, A. Bhargava, K. Lucaites, H. Solini, M. Nasiri, A. Robb, C. Pagano, B. Ullmer, et al. Empirically evaluating the effects of perceptual information channels on the size perception of tangibles in near-field virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 1–10. IEEE, 2021. 3

[12] 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. 2, 8

[13] E. Ebrahimi, S. V. Babu, C. C. Pagano, and S. Joerg. Towards a comparative evaluation of visually guided physical reach motions during 3d interactions in real and virtual environments. In *2016 IEEE symposium on 3D user interfaces (3DUI)*, pp. 237–238. IEEE, 2016. 7

[14] E. Ebrahimi, S. V. Babu, C. C. Pagano, and S. Jörg. An empirical evaluation of visuo-haptic feedback on physical reaching behaviors during 3d interaction in real and immersive virtual environments. *ACM Transactions on Applied Perception (TAP)*, 13(4):1–21, 2016. 7

[15] 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. 1, 8

[16] 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, 2016. 1, 7, 8

[17] F. Fontana and D. Rocchesso. Auditory distance perception in an acoustic pipe. *ACM Transactions on Applied Perception (TAP)*, 5(3):1–15, 2008. 2

[18] M. Frank and D. Perinovic. Matching auditory and visual room size, distance, and source orientation in virtual reality. In *Proceedings of the 17th International Audio Mostly Conference*, pp. 80–83, 2022. 5

[19] H. Gagnon, S. Creem-Regehr, and J. Stefanucci. Virtual room re-creation: A new measure of room size perception. In *ACM Symposium on Applied Perception 2021*, pp. 1–10, 2021. 2, 3, 5, 7, 8

[20] M. A. Goodale and A. D. Milner. Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25, 1992. 7

[21] L. T. Graham, S. D. Gosling, and C. K. Travis. The psychology of home environments: A call for research on residential space. *Perspectives on Psychological Science*, 10(3):346–356, 2015. 1

[22] S. Hameed, J. Pakarinen, K. Valde, and V. Pulkki. Psychoacoustic cues in room size perception. In *Audio Engineering Society Convention 116*. Audio Engineering Society, 2004. 2, 3

[23] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, vol. 52, pp. 139–183. Elsevier, 1988. 5

[24] D. Henry and T. Furness. Spatial perception in virtual environments: Evaluating an architectural application. In *Proceedings of IEEE virtual reality annual international symposium*, pp. 33–40. IEEE, 1993. 2

[25] M. Inui and T. Miyata. Spaciousness in interiors. *Lighting Research & Technology*, 5(2):103–111, 1973. 2

[26] N. Kaplanis and J. Van Velzen. Hearing through darkness: A study of perceptual auditory information in real rooms and its effect on space perception. *Proceedings of the Institute of Acoustics*, 34(4):114–125, 2012. 3, 8

[27] G. Kearney, M. Gorzel, H. Rice, and F. Boland. Distance perception in interactive virtual acoustic environments using first and higher order ambisonic sound fields. *Acta Acustica united with Acustica*, 98(1):61–71, 2012. 2

[28] J. W. Kelly, L. A. Cherep, B. Klesel, Z. D. Siegel, and S. George. Comparison of two methods for improving distance perception in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 15(2):1–11, 2018. 7

[29] J. W. Kelly, L. A. Cherep, and Z. D. Siegel. Perceived space in the htc vive. *ACM Transactions on Applied Perception (TAP)*, 15(1):1–16, 2017. 2

[30] K. Kohm, S. V. Babu, C. Pagano, and A. Robb. Objects may be farther than they appear: depth compression diminishes over time with repeated calibration in virtual reality. *IEEE transactions on visualization and computer graphics*, 28(11):3907–3916, 2022. 1

[31] A. J. Kolarik, B. C. Moore, P. Zahorik, S. Cirstea, and S. Pardhan. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics*, 78:373–395, 2016. 2

[32] G. E. Legge, R. Gage, Y. Baek, and T. M. Bochsler. Indoor spatial updating with reduced visual information. *PLoS One*, 11(3):e0150708, 2016. 2, 3

[33] W.-Y. Lin, R. Venkatakrishnan, R. Venkatakrishnan, C. Pagano, S. V. Babu, and W.-C. Lin. An empirical evaluation of the calibration of auditory distance perception under different levels of virtual environment visibilities. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2024. 1, 4, 8

[34] 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. 1, 2, 3, 4, 7, 8

[35] M. Loyola. The influence of the availability of visual cues on the accurate perception of spatial dimensions in architectural virtual environments. *Virtual Reality*, 22(3):235–243, 2018. 1, 2

[36] H.-J. Maempel and M. Jentsch. Auditory and visual contribution to egocentric distance and room size perception. *Building Acoustics*, 20(4):383–401, 2013. 2

[37] A. D. Milner and M. A. Goodale. Two visual systems re-viewed. *Neuropsychologia*, 46(3):774–785, 2008. 7

[38] 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. 1
[39] 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. 7
[40] C. C. Pagano and G. P. Bingham. Comparing measures of monocular distance perception: Verbal and reaching errors are not correlated. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4):1037, 1998. 7
[41] J. Pätynen, V. Pulkki, and T. Lokki. Anechoic recording system for symphony orchestra. *Acta Acustica united with Acustica*, 94(6):856–865, 2008. 3
[42] J. Ping, D. Weng, Y. Liu, and Y. Wang. Depth perception in shuffleboard: Depth cues effect on depth perception in virtual and augmented reality system. *Journal of the Society for Information Display*, 28(2):164–176, 2020. 2
[43] 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. 3
[44] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR)*, 46(2):1–40, 2013. 1, 2
[45] A. R. Richardson and D. Waller. The effect of feedback training on distance estimation in virtual environments. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 19(8):1089–1108, 2005. 2
[46] E. K. Sadalla and D. Oxley. The perception of room size: The rectangularity illusion. *Environment and Behavior*, 16(3):394–405, 1984. 1, 2
[47] C. S. Sahm, S. H. Creem-Regehr, W. B. Thompson, and P. Willemsen. Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM Transactions on Applied Perception (TAP)*, 2(1):35–45, 2005. 7
[48] A. Saulton, T. J. Dodds, J. Tesch, B. J. Mohler, and H. H. Bülthoff. The influence of shape and culture on visual volume perception of virtual rooms. In *Proceedings of the ACM Symposium on Applied Perception*, pp. 142–142, 2013. 2
[49] T. W. Schubert. A new conception of spatial presence: Once again, with feeling. *Communication Theory*, 19(2):161–187, 2009. 1
[50] G. Simpson, A. Sinnis-Bourozikas, M. Zhao, S. Aseeri, and V. Interrante. A virtual reality investigation of the impact of wallpaper pattern scale on qualitative spaciousness judgments and action-based measures of room size perception. In *Virtual Reality and Augmented Reality: 15th EuroVR International Conference, EuroVR 2018, London, UK, October 22–23, 2018, Proceedings 15*, pp. 161–176. Springer, 2018. 2, 3
[51] A. E. Stamps III. Effects of area, height, elongation, and color on perceived spaciousness. *Environment and Behavior*, 43(2):252–273, 2011. 2
[52] A. Tajadura-Jiménez, P. Larsson, A. Väljamäe, D. Västfjäll, and M. Kleiner. When room size matters: acoustic influences on emotional responses to sounds. *Emotion*, 10(3):416, 2010. 1
[53] E. M. Taranta, S. L. Koh, B. M. Williamson, K. P. Pfeil, C. R. Pittman, and J. J. LaViola. Pitch pipe: An automatic low-pass filter calibration technique for pointing tasks. *Proceedings of Graphics Interface 2019*, 2019. 8
[54] R.-K. Thériault, J. D. Manduca, and M. L. Perreault. Sex differences in innate and adaptive neural oscillatory patterns link resilience and susceptibility to chronic stress in rats. *Journal of Psychiatry and Neuroscience*, 46(2):E258–E270, 2021. 8
[55] M. Tolchinsky, R. Venkatakrishnan, R. Venkatakrishnan, C. C. Pagano, and S. V. Babu. Empirical evaluation of the effects of visuo-auditory perceptual information on head oriented tracking of dynamic objects in vr. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1074–1083. IEEE, 2023. 1
[56] A. Turner, J. Berry, and N. Holliman. Can the perception of depth in stereoscopic images be influenced by 3d sound? In *Stereoscopic Displays and Applications XXII*, vol. 7863, pp. 71–80. SPIE, 2011. 1, 7, 8
[57] C. Vienne, S. Masfrand, C. Bourdin, and J.-L. Vercher. Depth perception in virtual reality systems: effect of screen distance, environment richness and display factors. *IEEE Access*, 8:29099–29110, 2020. 2
[58] C. Von Castell, D. Oberfeld, and H. Hecht. The effect of furnishing on perceived spatial dimensions and spaciousness of interior space. *PloS One*, 9(11):e113267, 2014. 1, 2
[59] J. B. Wagman, K. Shockley, M. A. Riley, and M. Turvey. Attunement, calibration, and exploration in fast haptic perceptual learning. *Journal of Motor Behavior*, 33(4):323–327, 2001. 7
[60] Widex. Widex hearing test, 2021. 4
[61] R. Withagen and C. F. Michaels. Transfer of calibration in length perception by dynamic touch. *Perception & Psychophysics*, 66:1282–1292, 2004. 7
[62] M. Yadav, D. A. Cabrera, L. Miranda, W. L. Martens, D. Lee, and R. Collins. Investigating auditory room size perception with autophonic stimuli. In *Audio Engineering Society Convention 135*. Audio Engineering Society, 2013. 2, 3
[63] P. Zahorik. Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America*, 111(4):1832–1846, 2002. 2
[64] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst. Auditory distance perception in humans: A summary of past and present research. *ACTA Acustica united with Acustica*, 91(3):409–420, 2005. 2