The Effects of Laboratory Environment Type on Intermittent Sound Localization

Armisha Roberts, University of Florida,
Kyla McMullen, University of Florida

Virtual reality (VR) and augmented reality (AR) are gaining commercial popularity. 3D sound guidelines for AR and VR are derived from psychoacoustic experiments performed in contrived, sterile laboratory settings. Often, these settings are expensive, inaccessible, and unattainable for researchers. The feasibility of conducting psychoacoustic experiments outside the laboratory remains unclear. To investigate, we explore 3D sound localization experiments in-lab (IL) and out-of-the lab (OL). The IL study condition was conducted as a traditional psychoacoustic experiment in a soundproof booth. The OL condition occurred in a quiet environment of the participants' choosing, using commercial-grade headphones. Localization performance did not vary significantly for OL participants compared to the IL participants, with larger variation observed in the IL condition. Participants needed significantly more time to complete the experiment IL than OL. The results suggest that conducting headphone-based psychoacoustic experiments outside the laboratory is feasible if completion time is negligible.

INTRODUCTION

The commercial use of virtual and augmented reality is becoming more commonplace each day. Sound is a critical component of the virtual reality (VR) and augmented reality (AR) experience to enhance presence and interactivity. The sonic experience, especially 3D sound, impacts a user's situational awareness, focuses attention on visual cues, improves depth perception, increases navigation performance, improves the perception of visual stimuli, and provides complex information without overtaxing the visual system (Begault & Trejo, 2000).

Currently, the guidelines that researchers use to create realistic sounds for virtual and augmented environments are based on the results of psychoacoustic experiments conducted in heavily sanitized audio experiments. It is common practice to tightly control relevant experiment factors such as the sound, its characteristics, and the environment (often an anechoic or sound-treated room with one reflection). These experiments have given seminal insight into complex perceptual psychoacoustic phenomena.

Unfortunately, conducting experiments in a tightly controlled environment is not accessible for most researchers and participants. For example, a sound-treated room can cost between tens of thousands of dollars to one million dollars (USD). Barring price as an obstacle, physically traveling to a specialized, sound-treated research lab can be impossible due to many factors such as a global pandemic, lack of transportation, or power outage. Thus, more research should be conducted on the suitability of standard room environments on psychoacoustic experiment results. Removing the need for a specific testing environment can increase the amount and diversity (age, gender, background, etc.) of participants who can complete the experiment, thus making the experiment results more widely applicable.

There is limited research on the effects of the experiment environment on research performance. In other words, more research should investigate the strict necessity of sound-treated environments for accurate psychoacoustic perceptual experiments. Although, work has been performed by Talcott et al. to evaluate localization in a realistic environment, specifically military environments, this work focused on the use of hearing protection enhancement devices and not normal hearing localization (Talcott et al., 2012). The present work conducts a psychoacoustic, 3D audio localization experiment on participants in the lab (IL) and out of the lab (OL) using a normal, uninterrupted sound stimulus and more challenging sound sources containing various amounts of silence. The present work makes the case that headphone-delivered 3D audio experiments can be reliably performed under normal room conditions and do not need a sound-treated laboratory environment. The goal of this work is to:

1. Determine if the localization accuracy of OL participants is comparable to the accuracy of IL participants.
2. Determine if the amount of experiment time needed for OL participants is similar to OL participants.

BACKGROUND

Humans use sonic cues to determine the origin of sounds they hear in 3D space. This process is called localization. Localization was first investigated in the context of the Duplex Theory established by Rayleigh in 1907, which highlights the significance of the sonic cues, interaural time difference (ITD) or also referenced as interaural phase difference (IPD) within selected literature and interaural intensity difference (IID) or interaural level difference (ILD) (Macpherson & Middlebrooks, 2002; Middlebrooks, 2015; Middlebrooks & Green, 1991; Neuhoff, 2004). ITD is a sonic cue that highlights when sound hits one ear before the other. IID is a sonic cue that focuses on the intensity of a sound being louder at the closer ear versus that of the ear further away. These two cues influence how humans perceive the direction of low- and high-frequency sounds. For low-frequency sounds, below 1.5 kHz, localization is heavily dependent on the ITD as the sound wave will reach the closer ear before reaching the ear further away. On the other hand, for

high-frequency sounds above 1.5 kHz, localization is dependent on IID as the head casts a shadow, dampening the sound's intensity for the further away ear (Middlebrooks, 2015; Middlebrooks & Green, 1991; Neuhoff, 2004).

To render 3D audio, the sound must first be digitally filtered so the listener can perceive the specified direction of the sound. The rendering of 3D audio is done through Head-Related Transfer Functions (HRTFs), which are the Fourier transformation of Head-Related Impulse Responses (HRIRs). HRIRs are measured by placing small probe microphones in the left and right ear of a participant to record acoustic measurements of a sound being played at various locations in an anechoic chamber. These measurements are called binaural impulse responses (Mendonça, 2012).

METHODS

The purpose of this experiment was to examine the potential of conducting headphone-delivered 3D audio studies outside of the laboratory. To accomplish this task, we compared 3D audio localization results for in the lab (IL) and out-of-lab (OL) experiments. In the IL condition, participants were in a carefully controlled laboratory setting, using research-grade headphones, seated within a soundproof booth. In the OL experiments, participants completed the experiment in a quiet location of their choice. The purpose of this experiment is to develop a foundational understanding of how listeners localize standard and challenging intermittent sounds in a controlled and non-controlled setting. This experiment was reviewed and approved by the conducting institution's Institutional Review Board before data collection.

Participants

A between-subjects experimental design was conducted to evaluate performance data. In both experiments, 18-23-year-old novice listeners were chosen as participants. The reported gender of the IL participants are 6 Males, 5 Females, and 1 Non-binary person (total: 12); for the OL study, there were 29 Males, 14 Females, 2 Non-binary, and 2 Transmen (total: 47). All participants were students at the institution where this work was conducted. To be deemed eligible for the study, participants had to demonstrate, through a hearing screening, "normal hearing" by indicating the perception of tones that ranged from 125 Hz to 16 kHz. If a participant was not able to hear within the provided range, they did not participate in the experiment.

The hearing screening was facilitated using a standard protocol that assessed their detection of tones ranging from 125 Hz to 22 kHz. To assess high-frequency perception, tones were played at frequencies beginning at 22 kHz, sweeping down to 8 kHz. Then, to assess low-frequency sounds, tones from 125 Hz to 8 kHz were played. Participants were given the option to listen to these tones at varying levels from -5 dbHL to 70 dbHL. If a participant required 60 dbHL or higher to hear any of the low-frequency tones, they were also removed from the study. Only 1 participant was removed from each experiment condition group for not passing the hearing screening.

Stimulus

Spatialized white noise was used as the stimulus throughout each experiment. The white noise was spatialized in the 50 ear-level azimuth positions measured in the CIPIC HRTF database (Algazi et al., 2000), which are: 80°, 65°, 55° on both hemispheres as well as in 5° steps between ±45°. White noise was selected as the stimulus, as research has widely accepted that broadband sounds are the most optimal to localize on the horizontal plane. Narrowband sounds are not optimal for localization as particular frequencies lack interaural cues to direct a listener to the sound source (Stevens & Newman, 1936). Within the CIPIC HRTF database, Subject 15's HRTF was used throughout the experiment, as research has suggested a preference for HRTFs from the CIPIC database for non-individualized HRTF experiments (McMullen et al., 2012; Shukla et al., 2018).

In the IL experiment, all sounds were presented through Etymotic ER-1 in-ear headphones. These headphones are a high-definition, flat frequency response, research-grade product that is only found in laboratory settings. 3D sounds were rendered on an HP Pavilion Notebook laptop in a double-walled sound-treated booth, constructed by Technical Acoustics Inc, within the researchers' lab. In the OL experiments, participants used any headphones that were available to them. These types of headphones included in-ear, on-ear, and over-ear styles. Sounds were rendered using their preferred headphones in addition to their personal computing device. 3D sounds were pre-rendered and saved as audio files to avoid any perceptual challenges due to 3D sound rendering on a low-computation device.

In all aspects of the experiment, the stimulus was played for 5 seconds with no volume changes. However, the intermittency (amount of silence within a 5 second sound) was varied throughout the experiment to present more challenging localization trials. Additional details about each sound are further explained in the *Experiment* section.

Experimental Design

Regardless of the condition, participants conducted six experimental tasks: training, baseline assessment, training with feedback, then Trial 1(50 ms of intermittent silence), Trial 2 (100ms of intermittent silence), and Trial 3 (250ms of intermittent silence).

The IL participants first consented to the study, followed by a brief hearing screening to ensure they had normal hearing. Prior to beginning the hearing test, participants calibrated their headphones to match the sound level of rubbing their hands together in front of their faces. This calibration was performed to ensure the sound's volume would be played at a comfortable level. The OL participants followed a similar structure, except the OL participants provided their demographic information before the hearing test. Apart from this difference, all other components of the experiment were identical.

A between-subjects experimental design was employed to conduct the assessment. Forty-eight participants completed the OL experiment condition. One participant was removed due to not passing the hearing screening. 13 participants completed the IL experiment condition, and 1 participant was removed for failing the hearing screening. Because a singular participant from both the OL and IL study failed to pass the hearing screening, they did not perform the experiment and are not included in the participant count within the *Participants* section.

Procedure

First, training was employed to familiarize the participants with the sounds, procedure, and interface. In the training procedure, sounds were played in 15-degree increments from 0 to 359 on the horizontal plane. Sounds were played in 22 unique locations and were accompanied by visual feedback to help the participant learn the direction from which the sound was originating. Each sound played for 5 seconds. The participant was allowed to revisit any desired sounds to better train themselves on localization.

After training, a baseline assessment was conducted in which participants were asked to indicate the location of 50 randomly-presented continuous sound sources across all 50 eligible locations.

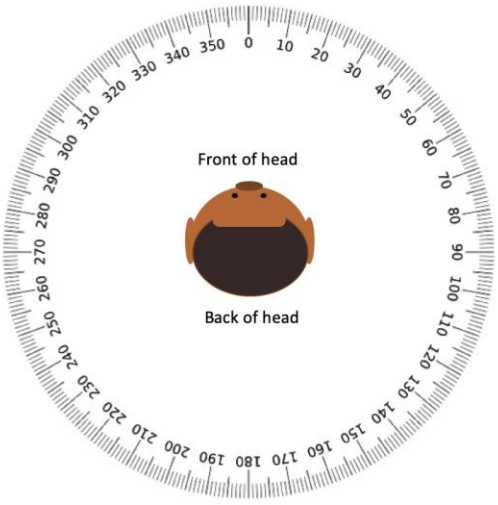

Figure 1: The visual interface that participants used to make their localization selection during the baseline assessment, training, and trials. Participants were asked to click the angle from which they perceived the sound to be emitting.

Because the experiment was conducted online, participants were presented with a top-down view of a head surrounded by 360 markers indicating the degrees around them (see Figure 1). Participants were instructed to imagine they were positioned at the center of the image, in the position of the head, and select the angle on the circle from which they perceived the sound to be originating. This visual interface was used in Baseline, training, and experimental trials.

After the baseline assessment, participants completed five blocks of training with feedback. Each block consisted of 10 localization tasks. The 3D sounds used in the five blocks had varying amounts of silence, ranging from 50ms, 100ms, 150ms, 200ms, and 250ms. Before localization, participants heard a 500ms reference sound inside their head. Afterward, 500ms of silence preceded the sound to be localized. The localization sound was played for 5 seconds. The reference sound, silence, localization sound pattern continued throughout the experiment. Participants could replay the localization sound as much as necessary. If the participant correctly answered 8 out of the ten localization questions correctly within the training block, they could proceed to the trial portion of the experiment. Participants were not allowed to begin the experiment trials until they correctly answered 8 out of the ten localization questions or completed all five training blocks.

The trial blocks consisted of 3 trials of 50 randomly spatialized, intermittent white noise. In Trial 1, participants localized 50 white noise sounds containing 50ms of silence. In Trial 2, participants localized 50 white noise sounds containing 100ms of silence. In Trial 3, participants localized 50 white noise sounds containing 250ms of silence. The trial blocks followed the same pattern as the training blocks with the reference sound, silence, and white noise localization sound. Participants could replay the localization sound as much as necessary.

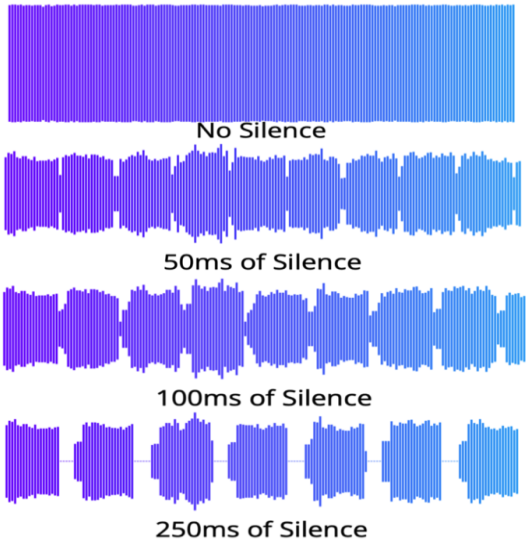

Figure 2: Visual representation of the sounds used in the Baseline and Trials 1-3 of the experiment with varying amounts of silence.

Evaluation Metrics

Similar to conventional localization studies, angular precision was used as the metric to assess accuracy. Within the context of this study, angular precision is defined as the difference between the actual angular location of the sound source and the angular location selected by the participant. To access angular accuracy selected within the visual interface, listeners were given a +/- 15° buffer for selecting the correct sound location. For instance, if a sound was played at 20°, but the participant

selected a value between 5° and 35°, their response was regarded as correct. This 15° accuracy is consistent with other experimental thresholds for trained listeners localizing 3D sounds on the horizontal plane (e.g., Larsen et al., 2013).

Front-back confusion is an ever-present challenge in localization tasks. Front-back confusion (FBC) occurs when sound sources that are on the "cone of confusion" are at an equal distance from the left and right ear. When this occurs, the sound sources share the same ITD and IID, which leads to confusion on where the sound source is truly coming from. Due to the sound sources sharing the same ITD and IID, it becomes difficult to distinguish if the sound is coming from in front or from behind. As is common in localization experiments, participants' responses were corrected for front-back confusion (e.g., Richter & Felds, 2016). The FBC rate was also used as an exclusion criterion for data analysis. If a participant had an FBC rate of 50% or higher, their data was removed from the analysis and was considered an outlier. Such a high FBC rate could indicate that the participant was inattentive, answering randomly, or was using insufficient headphones. As a result of this criteria, three participants were removed from the analysis of the IL condition, and 24 participants were removed from the analysis of the OL condition.

In addition to localization precision, the completion time was also assessed to determine if participants needed more time to complete the experiments IL vs. OL. The same participants that were removed from the IL and OL precision data analysis were removed from the time analysis.

Results

Angular Precision. A one-way analysis of variance (ANOVA) was performed to assess the difference in angular precision during each trial as an effect of the experimental setting (IL or OL). As noted between the IL and OL populations they have unequal sample sizes, however, an ANOVA can still be performed on this data. Results can be seen in Figure 3. The ANOVA compared the IL vs. OL angular accuracy for all participants for the Baseline (consistent sound), Trial 1, Trial 2, and Trial 3. For the Baseline condition, there was no significant difference in localization error IL as compared to OL ($F_{1,1599} = 0$, $p = .9446$). Similarly, there was no significant difference in performance between the IL and OL conditions for Trial 1 ($F_{1,1599} = .02$, $p = .8973$), Trial 2($F_{1,1599} = 2.23$, $p = .1354$), and Trial 3($F_{1,1599} = 3.51$, $p = .0611$).

Completion Time. A one-way analysis of variance (ANOVA) was performed to assess the difference in completion time for the entire experiment as an effect of the experimental setting (IL or OL). Results can be seen in Figure 4. The ANOVA compared the elapsed time for the IL participants and the OL participants. The average amount of time needed for the OL participants (11200.3 seconds) was significantly higher than the completion time needed for the IL participants (7261.5 seconds) ($F_{1,29} = 6.39$, $p < 0.05$). In addition, calculating the Pearson's correlation between the duration and accuracy of the participants' performance yielded weak results for both IL and OL. The IL correlation coefficients for the baseline trial, trial 1, trial 2, and trial 3 were -.27, .04, .18, and .09 respectively. The OL correlation coefficients for the baseline trial, trial 1, trial 2, and trial 3 were -.45, -.33, .14, and -.05 respectively.

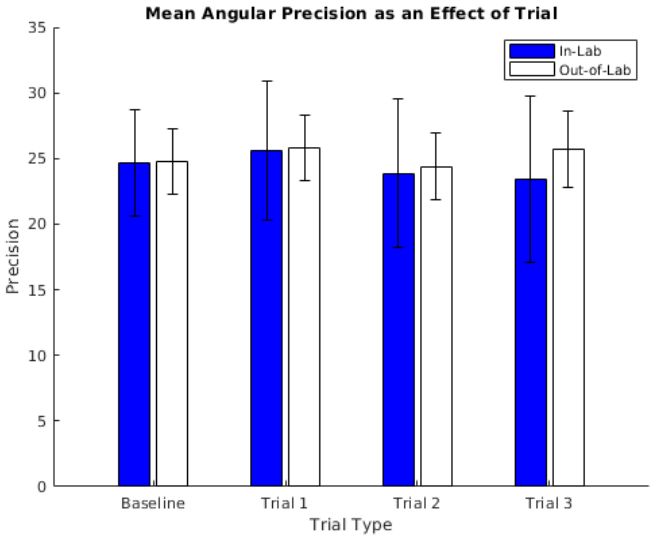

Figure 3: The mean angular precision based on the associated trial for in-lab and out-of-lab participants. The x-axis denotes the trial. The y-axis represents the mean precision, and the error bars represent the mean standard error.

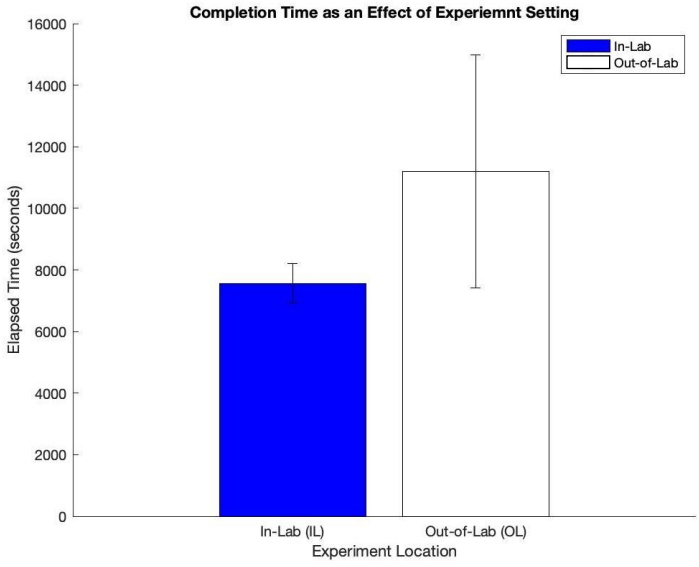

Figure 4: The difference in completion time for in-lab and out-of-lab participants. The x-axis denotes the experiment location. The y-axis represents the elapsed time in seconds, and the error bars represent the mean standard error.

DISCUSSION

The goal of the present work was to determine the viability of conducting reliable headphone-based 3D audio psychoacoustic experiments outside of the traditional laboratory setting. This

work could open the door to greater experimental possibilities. Psychoacoustic experiments could be conducted by more researchers. Greater amounts of people, as well as people from varying backgrounds, would be able to participate in psychoacoustic experiments, thus allowing the results to be relevant to and representative of a wider audience.

The present work explored a series of experiments to compare participants' angular precision during localization experiments using sounds with varying degrees of silence, presenting a challenge during in-lab (IL) and out-of-lab (OL) experimental conditions. It was observed that there was no significant difference in response precision during localization for the Baseline and trials between the IL and OL participants. This result suggests that headphone-based 3D audio psychoacoustic experiments can be reliably performed outside of the lab with similar accuracy as being performed in the lab. The mean angular precision were between 20-26°, which is similar to that reported in other seminal studies of novice listeners localizing headphone-based 3D sounds (e.g., Wightman & Kistler, 1989).

The present work also assessed the effects that the experimental conditions posed on the experiment's overall completion time. It was observed that the participants that conducted the experiment outside of the lab needed significantly more time and also had a higher amount of variability in responses. This finding suggests that out-of-lab participants should be allocated significantly more time to complete the experiment. One explanation for this significance is that the OL participants completed the experiment in a quiet and comfortable environment of their choosing. They may have paused and resumed work many times due to their environment. In the IL case, participants were in a research lab with few distractions, so the experimental condition did not take as long.

It is also vital to mention the potential experimental bias that may have taken place during this study. Due to the participants being recruited from a course, the sampling population was rather narrow and does not depict the broader population of individuals with normal hearing. The last form of bias to take into consideration is response bias. Participants were incentivized to complete the study; however, the researcher cannot control the participants' accuracy in reporting.

In the IL condition, three participants were removed because of high front-back confusion rates. In the OL condition, 24 participants were removed from the analysis. This observation suggests that even though OL experiments can produce similar perceptual results as IL experiments, many more participants will need to be excluded from the experiment data analysis.

BIBLIOGRAPHY

Algazi, V. R., Duda, R. O., Thompson, D. M., & Avendano, C. (2001, October). The CIPIC HRTF database. *In Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics* (Cat. No. 01TH8575) (pp. 99-102). IEEE.

Larsen, C. H., Lauritsen, D. S., Larsen, J. J., Pilgaard, M., & Madsen, J. B. (2013, September). Differences in human audio localization performance between a HRTF- and a non-HRTF audio system. I*n Proceedings of the 8th Audio Mostly conference* (pp. 1-8)

Macpherson, E. A., & Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America*, *111*(5), 2219. https://doi.org/10.1121/1.1471898

McMullen, K., Roginska, A., & Wakefield, G. H. (2012). *Subjective Selection of Head-Related Transfer Functions (HRTFs) based on Spectral Coloration and Interaural Time Differences (ITD) Cues*. 10.

Mendonça, C. (2012). *On the improvement of localization accuracy with non-individualized HRTF-based sounds*. 12.

Middlebrooks, J. C. (2015). Sound localization. In *Handbook of Clinical Neurology* (Vol. 129, pp. 99–116). Elsevier. https://doi.org/10.1016/B978-0-444-62630-1.00006-8

Middlebrooks, J. C., & Green, D. M. (1991). *Sound Localization by Human Listeners*. 25.

Neuhoff, J. G. (2004). Auditory Motion and Localization. In *Ecological Psychoacoustics* (pp. 87–111). Elsevier. https://doi.org/10.1016/B978-012515851-0/50005-9

Richter, J. G., & Fels, J. (2016). Evaluation of localization accuracy of static sources using HRTFs from a fast measurement system. *Acta Acustica United With Acustica*, 102(4), 763-771.

Shukla, R., Stewart, R., Roginska, A., & Sandler, M. (2018). *User selection of optimal HRTF sets via holistic comparative evaluation*. 10.

Stevens, S. S., & Newman, E. B. (1936). The Localization of Actual Sources of Sound. *The American Journal of Psychology*, *48*(2), 297. https://doi.org/10.2307/1415748

Talcott, K. A., Casali, J. G., Keady, J. P., & Killion, M. C. (2012). Azimuthal auditory localization of gunshots in a realistic field environment: effects of open-ear versus hearing protection-enhancement devices (HPEDs), military vehicle noise, and hearing impairment. *International journal of audiology*, *51*(sup1), S20-S30.

Wightman, F. L., and Kistler, D. J. (1989). Headphone simulation of free-field listening. II: Psychophysical validation. J*ournal of the Acoustical Society of America*, 85, 868–878.