

# Impact of wearing a head-mounted display on localization accuracy of real sound sources

David Poirier-Quinot  and Martin S. Lawless\*

Sorbonne Université, CNRS, UMR 7190, Institut Jean Le Rond d'Alembert, 4 place Jussieu, Paris, France

Received 14 April 2022, Accepted 21 November 2022

**Abstract** – For augmented reality experiences, users wear head-mounted displays (HMD) while listening to real and virtual sound sources. This paper assesses the impact of wearing an HMD on localization accuracy of real sources. Eighteen blindfolded participants completed a localization task on 32 loudspeakers while wearing either no HMD, a bulky visor HMD, or a glass visor HMD. Results demonstrate that the HMDs had a significantly impact on participants' localization performance, increasing local great circle angle error by 0.9°, and that the glass visor HMD demonstrably increased the rate of up-down confusions in the responses by 0.9–1.1%. These results suggest that wearing an HMD has a sufficiently small impact on real source localization that it can safely be considered as an HMD-free condition in most but the most demanding AR auditory localization studies.

**Keywords:** Binaural, Localization, Head-mounted display

## 1 Introduction

Binaural hearing refers to how spatial audio is perceived with two ears. As an acoustic wave travels towards the eardrums, it interacts with the listener's head, torso, and pinna, causing interaural timing and level differences (ITDs and ILDs, respectively) between the left and right ears as well as spectral distortions unique to each ear. The human brain exploits these interaural differences and spectral distortions, specific to each individual's morphology [1], to infer the position of a source in space. Previous studies have shown that changes in morphology, such as wearing earmolds [2] or headgear [3], can have a significant impact on the ability to localize sounds in space. The advent of virtual and augmented reality (VR and AR), frequently requiring users to wear a head-mounted display (HMD) while simultaneously listening to real and virtual sound sources, raises the question of how much wearing an HMD impacts auditory perception of the real sources.

Wearing an HMD has been shown to impact head-related transfer functions (HRTFs), which are filters that describe the interaural differences and spectral distortions caused by a person's morphology. Previous works [4–8] have measured and compared the HRTFs of humans and acoustic manikins with and without HMDs on grid densities ranging from 6 points on a horizontal semicircle to 2702 points homogeneously distributed around the head. These studies reported that HMDs induced slight differences in

the 1–5 kHz frequency range and more pronounced differences in 5–16 kHz, predominately in the contralateral ear. The magnitude of the frequency spectrum was reported to be affected mainly in spatial locations between  $-120^\circ$  and  $-30^\circ$  in azimuth and  $-60^\circ$  and  $60^\circ$  in elevation in the contralateral ear, for both human and manikin HRTFs [4], similar to the changes induced by wearing a baseball cap [8]. Based on their objective impact on the HRTF, HMD designs may be sorted into two categories: bulky visors (e.g. Oculus Quest 1 and HTC Vive) and glass visors (e.g. Microsoft HoloLens 1 and MagicLeap). Bulky visor HMDs were reported to affect ITDs and ILDs for frontal directions [8], as well as ILD values at  $-60^\circ$  and  $-120^\circ$  in azimuth [7]. Glass visors, the HoloLens 1 in particular, created reflections in the signals that led to additional peaks in the mid-frequency range of the spectrum (3–7 kHz) at  $-120^\circ$  in azimuth [7].

Other works have examined the impact of HMDs on the perceived quality of HRTFs. These studies conducted multiple-stimuli hidden anchor and reference (MUSHRA) tests to evaluate differences in timbre and localization quality [4, 7, 9]. The reported perceptual degradation agreed with the objectively measured differences [7], namely that the impact of the HMD on the listener's own HRTF was perceivable, but not as noticeable as differences between the listener's HRTF and a generic one [4]. According to Lladó et al. [9], bulky visors HMDs can be expected to slightly alter the timbre of sources directly in front of the listener, i.e. directly hitting the main bulk of those HMDs.

\*Corresponding author: [davipoir@dalembert.upmc.fr](mailto:davipoir@dalembert.upmc.fr)

Only two studies to date were found to have assessed the impact of wearing an HMD on the localization accuracy of real sound sources [3, 9]. In Lladó et al. [9], participants performed two separate localization tasks with 18 sources positions in the horizontal plane and 14 in the median plane, respectively, with and without wearing a Quest 2 HMD (bulky visor HMD). Results suggested that the HMD had no impact on azimuth and polar errors in the frontal hemisphere, and no impact on quadrant errors (response 90° away from target position) overall. In Ahrens et al. [3], participants performed a localization task with 27 source positions, reported in Figure 1a, located in the frontal hemisphere and limited elevations (3 levels between -30 and 30°) with and without wearing a bulky visor HMD, specifically the HTC Vive 1. Results suggested that the HMD had no impact on azimuth localization at 0° elevation and a slight impact on elevation localization, increasing the elevation error by 1.8° compared to the No HMD condition. While HMDs appeared to cause coloration that affects the perceptual quality of stimuli, they may only have a minor impact on localization performance. However, due to the limitation of the source positions in the previous studies, further evaluation is required to (1) characterize the impact of wearing an HMD on localization of sources located on the whole sphere, including the rear hemisphere, as well as (2) to assess how the overall localization accuracy is impacted by various types of HMD designs.

The current study aims to further investigate the impact of HMDs on localization accuracy by including source locations over the whole sphere, as well as comparing the two main types of HMD designs. A localization test was conducted on 32 real source locations positioned around listeners with a bulky visor HMD, a glass visor HMD, and without HMD. Localization performance was assessed by computing the lateral, polar, and great circle angle errors between 32 real loudspeaker positions and the locations indicated by the participants. The percentage of localization reversals were also evaluated to determine if the HMDs induced front-back or up-down confusions. Based on the previous research, it was hypothesized that the HMDs would have little effect on the localization performance of the participants, and that observed significant differences would only concern perceived polar angle.

## 2 Materials and methods

Participants performed a localization task in 3 different conditions: once with each HMD under consideration and once without wearing an HMD for control (denoted the *No HMD* condition). The two HMDs used in the study, a Meta Quest 2 and a Microsoft HoloLens 1, were selected as representative devices of bulky and glass visors, respectively. This choice was guided by an informal comparison of the HMD designs available at the time of writing, all resembling, or fitting within the shape of, one of the two headsets. The Quest 2 has a large display that sits over the eyes, but otherwise has a low profile on the rest of the head. The display measures 16.5 cm across, is 9 cm tall,

and protrudes by 7.5 cm off the user's face. The HoloLens 1 has a slightly smaller profile on the face (15 cm across, 7 cm height, and 6 cm depth off of the face, ignoring curvature), but has a 3 cm tall headband that protrudes approximately 3 cm over the ears.

### 2.1 Participants

Eighteen participants volunteered to partake in the study. All but two of the participants were right-handed, the others being left-handed and ambidextrous respectively. Only one of the participants reported that they had been diagnosed with mild hearing loss. This participant was still included because the study was interested in comparing relative differences in accuracy between the three different conditions and the participant demonstrated comparable localization accuracy to the other participants without reported hearing losses.

Before the test session began, the participants completed a questionnaire asking about their previous experience with localization tasks. Eight of the participants had never participated in a localization experiment before, seven had participated in at least one, and three had participated in more than three localization experiments. It is important to note that the three participants that had participated in several experiments were previously familiar with the locations of the loudspeakers in the room, while the others were not. Thus, the participants were organized into these three categories of localization task experience (novice, intermediate, and expert) for the analyses.

Two additional participants (1 female, 1 male) were excluded from the analysis because their responses were only located in the frontal hemisphere for all three sets, including the No HMD condition, despite 14 of the 32 loudspeakers being located behind the participants. Both of these participants had no reported hearing losses and had never participated in a localization test before.

The order in which the HMD conditions were presented was counterbalanced across participants and consisted of the permutation of the three HMD conditions. Each participant was pseudo-randomly sorted into one of six equally-sized groups that determined the order in which they received the different conditions to complete the localization task.

### 2.2 Localization task description

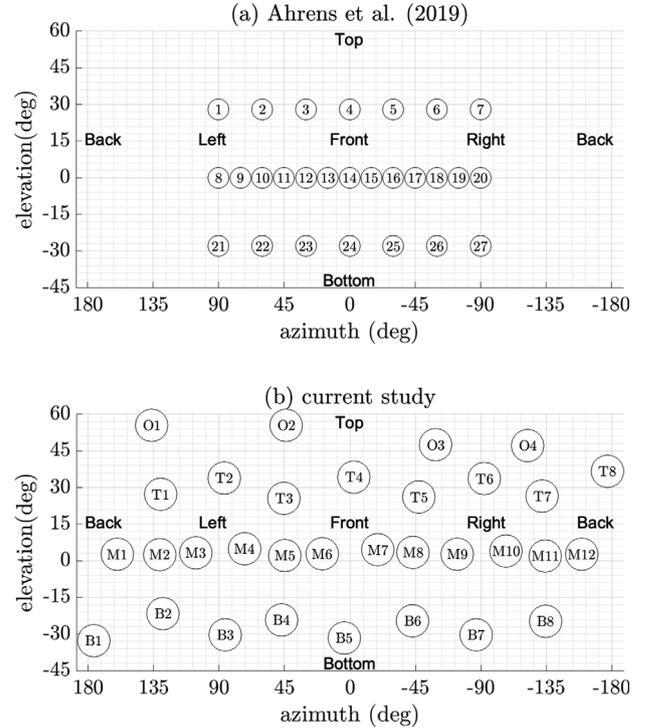
The localization task was performed in an acoustically damped room with a mid-frequency reverberation time  $T_{30,1000\text{Hz}}$  of 0.12 s and an ambient noise level below 30 dBA. Thirty-two loudspeakers were placed at  $2.4 \pm 0.5$  m from participants, on the surface of a rectangular cuboid centered on the room at three vertical levels from the ground, spanning elevations from -32° to 56°. The bottom, middle, and top row of the grid contained 8, 12, and 8 speakers at approximate heights of 0.1, 1.4, and 2.6 m respectively, spanning elevations from -32° to 36°. These speakers, relatively uniformly distributed around each row, will accordingly be referred to with speaker IDs

consisting of their row and number as B1–B8, M1–M12, and T1–T8. An additional four overhead speakers (O1–O4) were placed at a height of 2.6 m and have corresponding elevation angles between  $46^\circ$  and  $56^\circ$ . Exact speaker positions are reported in Figure 1b. Each speaker was individually equalized using a 2048-point finite impulse response filter that was calculated with measurements for each speaker from their respective locations to a receiver at the approximate listener head position.

The experiment was conducted with a Cycling’74 Max v8.0.5 patch, which controlled the stimulus presentations from the loudspeakers via RME Madiface and Fireface 800 sound cards. The patch also logged the positions of the center of the participants’ head and that of a tracked controller held by the participants, used to indicate perceived source locations. Tracking was performed using an array of 10 Optitrack Flex 13 cameras distributed around the room. The stimuli were sequences of three bursts of white noise, 40 ms each with a 4 ms cosine-squared onset/offset ramp separated by a 30 ms silence, for a total duration of 180 ms [10]. Full spectrum burst sounds were chosen to optimize localization ability [11], while the bursts had a short duration to limit the use of head-movement for localization. To avoid participants becoming familiar with the noise signal, 8 different pre-generated bursts were used, as well as random gains in a  $\pm 3$  dB range applied to each stimulus. For the localization task, the stimuli were presented from one speaker at a time and the participants were prompted to indicate the location of the noise. Each speaker position was repeated three times within a set for a total of 96 stimulus presentations per HMD condition. To reduce systematic procedural learning, the participants were not provided feedback on the accuracy of their responses.

Wearing the Hololens 1 partially blocks users field of view, while they cannot see the room at all with the Quest 2. To address this potential source of variability, the participants were blindfolded for each of the conditions. After completing the questionnaire, the blindfold was provided before participants were led into the testing room such that they could not see the locations of the speakers prior to the testing. Participants were placed on a stool in the center of the room and the height of the stool was adjusted such that the center of each participant’s head was at a height of 1.3 m. Subsequently, each participant performed a short, guided training exercise to familiarize them with the experimental procedure and to ensure that they were using the tracked controller correctly.

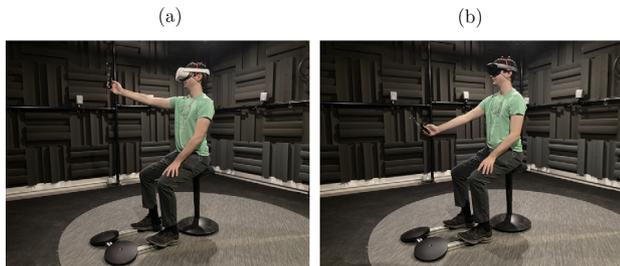
Participants were required to sit in the center of the array, facing forward to receive a stimulus presentation. Since they were blindfolded and were allowed to orient themselves to indicate source locations, tactile markers were placed on the ground so that the participants could find the forward position with their feet. During the localization task, the experimenter would help participants who needed assistance finding the correct position using a push-to-talk microphone connected to speakers in the array. A 100 ms, 440 Hz pure-tone pulse was used to inform participants that they were in the correct location, followed by 1 s of silence



**Figure 1.** Audio source positions tested in (a) Ahrens et al. [3] and (b) current study. Coordinates are relative to participant’s head position. Orientation labels “Back”, “Left”, etc. have been moved up from  $0^\circ$  elevation for readability.

before the stimulus presentation. Participants were instructed to indicate the origin direction of the stimulus by holding out the tracked controller at arms-length and imagining a line connecting the center of their head at the moment of the response, their thumb on the controller button, and the source of the noise. With the arm fully out-stretched and the pad of thumb being the reference point for the participants, hand rotation would not affect the response location. Participants were given the controller prior to being blindfolded and asked to point at specific objects in the pre-testing room to make sure they understood how the reporting method worked and ensure accuracy in the responses. This reporting method was used as it is consistent, accurate, and is not susceptible to pointing bias caused by participants not being able to see the tracked controller orientation while wearing the blindfold [3, 12]. The participants were allowed to turn to indicate source positions behind them and were encouraged to switch hands, to ensure that the tracked controller was always held at arms length to improve pointing accuracy. The experimental setup for both Quest 2 and Hololens 1 conditions is illustrated in Figure 2.

The participants completed a total of three sets of the localization task, once for each HMD condition. Each set was completed in approximately 10 min. Participants were given the option to take a 5 min break between sets to reduce the effect of test fatigue. Participants who did not take a break after the first set were required to do so after the second set.



**Figure 2.** Experimental setup during the (a) Quest 2 and (b) Hololens 1 conditions.

### 2.3 Data analysis

Localization accuracy was assessed based on 3 angular error metrics calculated between the actual source locations and participants responses, as well as the rate of localization reversal errors. *Great circle angle error* was calculated as the central angle of the two vectors originating from the center of the listener’s head during the response aimed towards the source and the indicated locations, respectively. Because great circle angle only provides a magnitude of error, the *lateral angle error* and *polar angle error* were computed, using the interaural coordinate system. The lateral and polar errors are the signed difference between the lateral and polar angles of the target and the participant response. Since polar angle error becomes distorted at the poles of the coordinate system [13], a  $0.5 \times (\cos(2 \times \text{lateral\_angle}_{\text{source}}) + 1)$  weight was applied to the polar error to compensate for the compression [14]. Absolute angle error values are also used in conjunction with signed errors, when the analysis focuses on error magnitude rather than localization bias. The term “absolute” or “abs.” is systematically used when referring to the former.

Since localization reversals inflate angular metric values, *local* angular metrics were also calculated and analyzed. Local great circle, lateral, and polar angle errors were computed for responses that fell within a  $45^\circ$  cone around the actual source location to distinguish local accuracy errors from those resulting from localization reversals. Analyses of variances (ANOVAs) [15] were conducted for each of the dependent variables of mean global and local great circle angle, lateral, and polar errors in R [16], to assess the effect of the different factors of HMD condition, loud-speaker position ID, localization task experience, and the first-order interaction terms between them.

Additionally, the number of localization reversals errors, i.e. front–back or up–down confusion rates, were calculated and compared across the aforementioned factors based on the polar angle classification scheme described in Zagala et al. [17]. As for the polar error, the scheme was altered to avoid compression at the poles inflating up–down and front–back rates [14, 18]: excluding targets with absolute lateral angle above  $67.5^\circ$  (within a  $45^\circ$  cone from the poles) from the analysis. Generalized linear mixed models (GLMMs), constructed as repeated measures logistic regressions [15], were used to evaluate the effects of HMD, localization task experience, and the interaction term on the

percentage of reversal errors. Likelihood ratio tests were performed to determine the significance of each factor by comparing the goodness of fit between the full models and models with single-term deletions.

Since previous studies have shown that procedural training may occur over the course of a localization test [14], additional statistical tests were conducted to examine effects of learning/fatigue on the angular error metrics and localization reversal errors. To determine if learning or fatigue occurred, that is if the participants overall performance improved or worsened over the course of the experiment, ANOVA and GLMM tests were conducted with set number (1, 2, or 3) as a factor of the angular metrics and localization reversals, respectively.

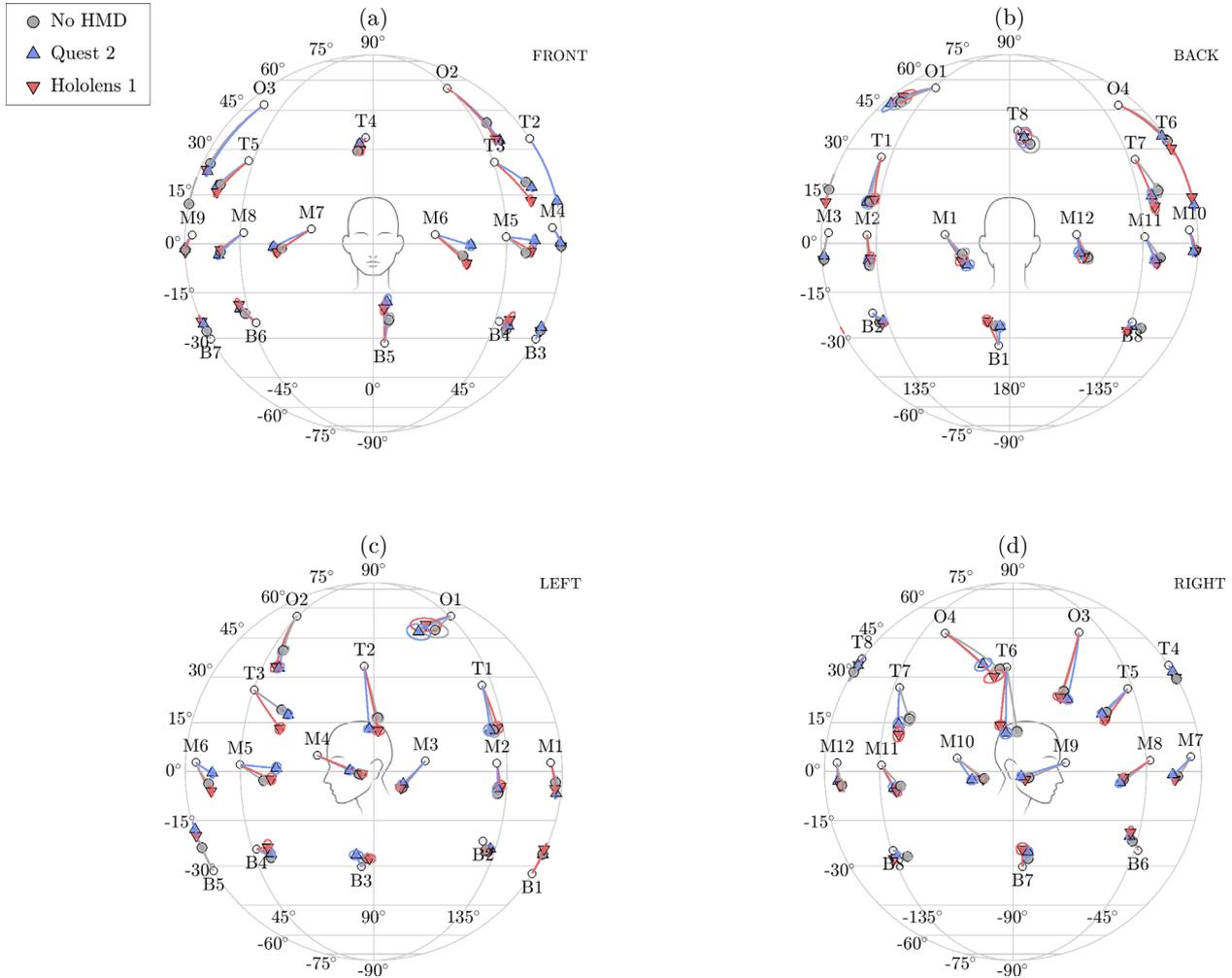
For all of the tests, statistical significance was determined for  $p$ -values below a 0.05 threshold. The notation  $p < \epsilon$  is adopted to indicate  $p$ -values below  $10^{-3}$ . Post-hoc pairwise comparisons for significant factors were made with Tukey–Kramer adjusted  $p$ -values.

## 3 Results and discussion

### 3.1 Analysis of the No HMD condition

The analysis of the No HMD condition was conducted first to serve as a baseline for later comparison with both HMD conditions. The average participant response locations were first plotted on a sphere to observe overall trends (see Fig. 3). The responses appeared laterally dilated, meaning that the responses were shifted towards the interaural axis compared to the actual speaker locations. Additionally, the perceived locations of the top and bottom speakers tended to be compressed towards the horizontal plane. Responses associated to sources in the middle row, which were approximately 0.1 m above the horizontal plane of each participant, had negative polar errors on average, which may have been caused by the blindfolded participants defaulting their responses towards the horizontal plane or a pointing bias caused by the response procedure. Interestingly, there were no apparent differences between sources in the front versus back hemispheres, even though participants had to turn around to point towards the back hemisphere source locations.

Based on participant responses for the No HMD condition session, Figure 4a shows the great circle angle error for all 32 speaker positions on a Voronoi diagram. The diagram partitions the azimuth/elevation plane into regions, referred to as cells, that contains all the points of the plane that are closer to a particular speaker than to any other. Separate repeated measures ANOVAs were conducted on all angular metrics to examine the impact of speaker position (i.e. speaker ID). The speaker position had a significant impact, on all angular metrics: great circle angle error ( $F = 13.0$ ,  $p < \epsilon$ ), lateral error ( $F = 19.9$ ,  $p < \epsilon$ ), and polar error ( $F = 7.00$ ,  $p < \epsilon$ ). Pairwise comparisons across speakers indicated that participants were less accurate for overhead sources (speakers O1–O4). The low performance for these overhead sources may be attributed to



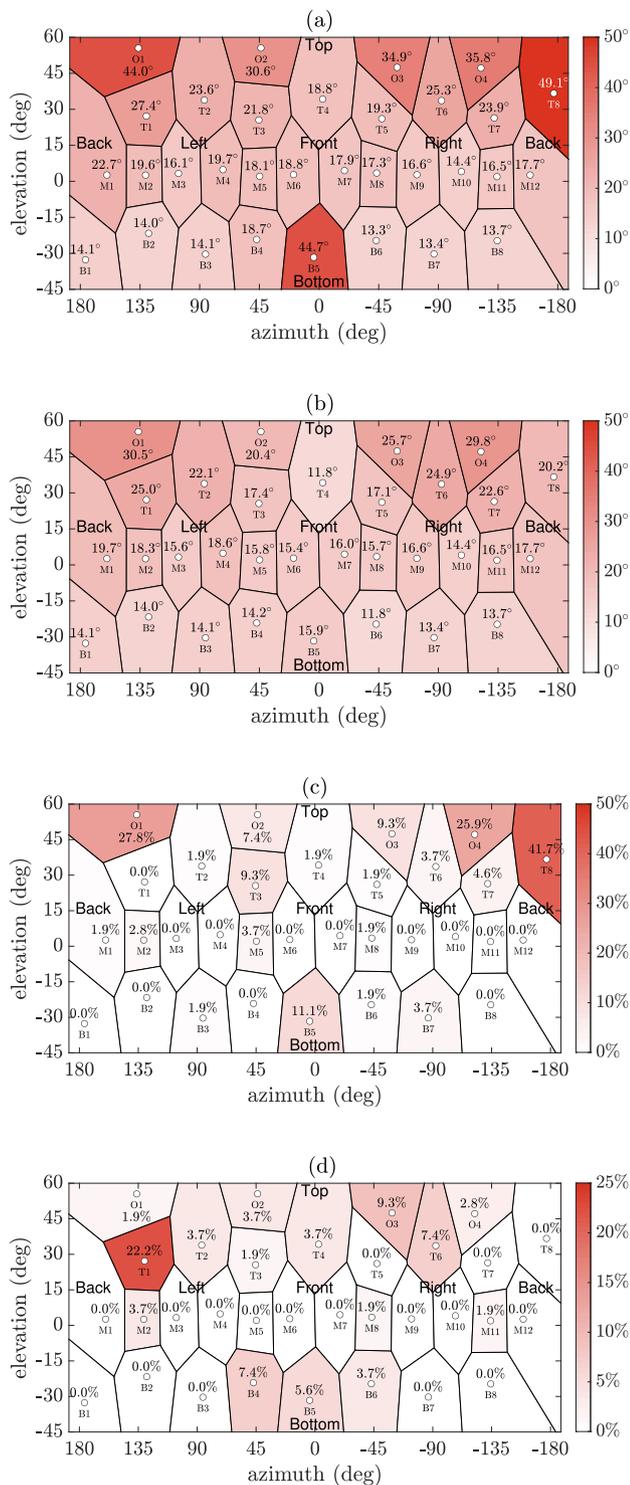
**Figure 3.** Average participant response locations for each HMD condition across speaker positions in the (a) front, (b) back, (c) left, and (d) right hemispheres. Only responses that fell within a  $45^\circ$  aperture cone around the actual source location are considered, discarding confusions to focus on local accuracy errors. Ellipses represent the standard error of the major and minor axes of the data variance obtained using the Kent distribution, after Leong and Carlile [13].

low perceptual accuracy in that region, as reported in Blauert [1] with a single overhead source location.

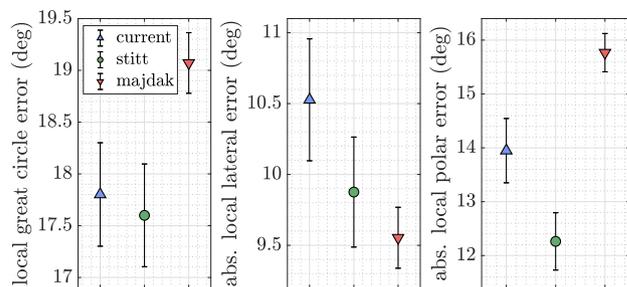
More surprisingly, poor localization performance was also observed for speakers B5, T1, and T8. When localization reversal errors were removed from the analysis, the local great circle angle of these speakers (Fig. 4b) greatly decreased, which means that the participants exhibited a high amount of front-back (Fig. 4c) and up-down (Fig. 4d) reversals for these particular speaker locations. The cause of the large confusion rates for these speakers may be due to effects of the room on localization performance. As shown in Figure 2, the experiment was conducted in an acoustically-damped room that had a reverberation time ( $T_{30}$ ) of 0.12 s for the 1000 Hz octave band. As there was a 10 cm gap in acoustic foam on the walls near these speakers, in addition to the carpet being less absorbing than the acoustic foam treatment on the walls and ceiling surfaces, the poor localization performance may be due to reflections from the surfaces at these specific

positions [19]. Impulse response measurements conducted from all of the speaker positions demonstrated a combing effect due to the proximity of the speakers to the surfaces of the room that was more pronounced for the speakers not near absorbing materials.

The results from the No HMD condition were compared to those of two other studies on source localization to validate the current experiment. In Majdak et al. [14], participants localized auditory stimuli produced with individual HRTFs, using either their head or their hand to indicate a direction of arrival in a virtual environment. Stitt et al. [20] also presented auditory signals to the participants using their individual HRTFs and used a reporting method similar to that used in the present study, except that the participants were not blindfolded. In both of these studies, participants were trained with over several hundreds of stimuli presentations until their performance was no longer improving, which nominally occurred after 1000 trials.



**Figure 4.** Localization performance across speaker positions, depicting (a) great circle angle error, (b) local great circle angle error, (c) front-back confusion rate and (d) up-down confusion rate for the No HMD condition. Metric values are reported above each speaker position (white circles), around which Voronoi cells have been drawn to simplify results interpretation. The apparent cell structure asymmetry is a result of the projection mapping inducing distortion at the poles (near  $\pm 90^\circ$  elevation).



**Figure 5.** Comparison of participants performance during the No HMD condition and the performance plateaus reported in Stitt et al. [20] and Majdak et al. [14]. For the former, reported results are those of participants trained with individual HRTFs. For the later, reported results are those of their experiment 2, aggregating head and hand reporting methods. For both, only the trials after participants reached a performance plateau were considered, i.e. discarding the first 1000 trials. The right-most figure reports absolute polar error values weighted according to the scheme presented in Section 2.3.

To compare the results to the current study, the angular errors from the previous studies were computed after discarding the first 1000 trials of each participant to evaluate their best performance (see Fig. 5). Localization performance during the No HMD condition of the present study was comparable to those of Majdak et al. [14] and Stitt et al. [20]. The average local great circle angle error in the No HMD condition was similar to that of Majdak et al. [14] at  $17.8^\circ$  and  $17.6^\circ$  respectively, while Stitt et al. [20] reported a higher amount of error at  $19.1^\circ$ . The absolute local lateral angle error in the previous studies were  $0.6^\circ$ – $0.9^\circ$  lower than the current study, which is not a significant amount. Last, the absolute weighted local polar angular error of the current study was  $1.3^\circ$  higher than reported in Majdak et al. [14] and  $1.9^\circ$  lower than reported in Stitt et al. [20]. These comparisons suggest that the current localization task and reporting method yield similar errors and uncertainties as those reported in the literature.

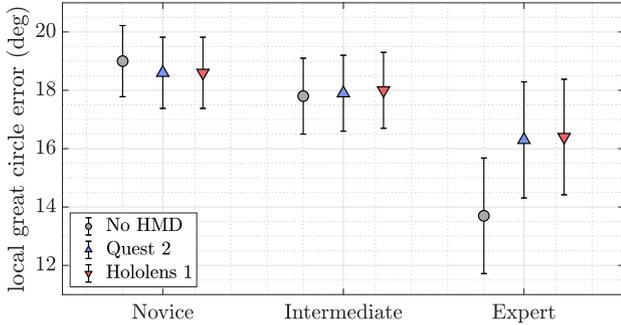
### 3.2 Impact of HMD on angular metrics

The impact of HMDs on the localization performance of the participants was first assessed on the angular metrics of local great circle angle, lateral angle, and weighted-polar angle errors. Repeated measures ANOVA tests were conducted with factors of HMD condition, speaker position ID, and localization task experience, as well as the first-order interaction terms of these main factors. As expected from the analysis on the No HMD condition, speaker ID had a significant impact on all these metrics ( $p < \epsilon$  for every ANOVA).

A summary of the results may be found in Table 1 that contains the mean values for the local angular error metrics. The main effect of HMD was significant for local great circle angle error ( $F = 3.91$ ,  $p = 0.02$ ) and local lateral angle error ( $F = 3.21$ ,  $p = 0.04$ ), but not for local weighted-polar angle

**Table 1.** Summary of localization performance for the three HMD conditions detailing the Tukey-adjusted local angular metrics and localization reversal rates. Reported  $\pm$  values for angular metrics are standard errors. Reported  $\pm$  values for localization reversal rates are lower and upper 95% confidence intervals.

Condition	Local angular errors ( $^{\circ}$ )			Localization reversal rates (%)	
	Great circle	Lateral	Weighted-polar	Front-back	Up-down
No HMD	$16.8 \pm 0.9$	$-2.0 \pm 0.7$	$0.0 \pm 0.7$	4.3 [3.0, 6.1]	1.5 [0.8, 2.9]
Quest 2	$17.6 \pm 0.9$	$-0.8 \pm 0.7$	$0.6 \pm 0.7$	4.5 [3.1, 6.3]	1.3 [0.7, 2.4]
Hololens 1	$17.7 \pm 0.9$	$-1.5 \pm 0.7$	$-0.2 \pm 0.7$	5.4 [3.8, 7.6]	2.4 [1.3, 4.4]

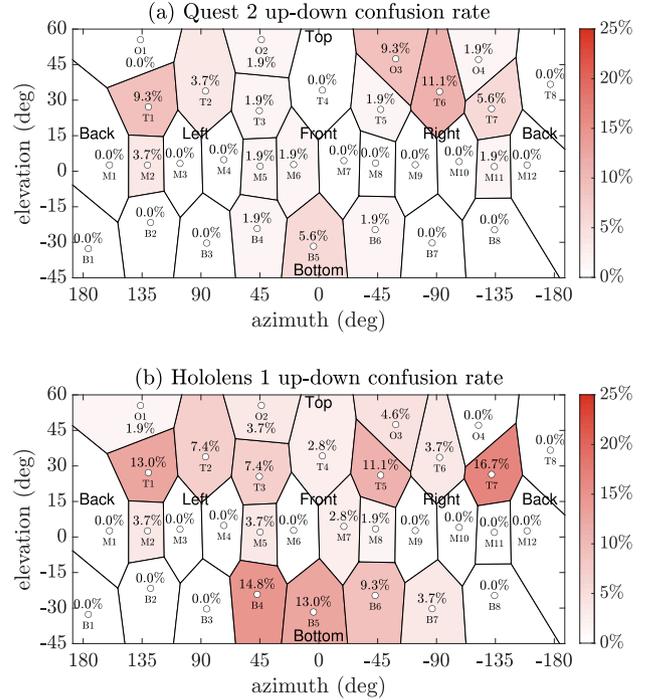


**Figure 6.** Local great circle angle error across HMD conditions and participants self-reported localization expertise.

error ( $F = 0.905$ ,  $p = 0.4$ ). For great circle angle error, a post-hoc pair-wise comparison test revealed that the participants exhibited a higher amount of great circle angle error of  $0.9^{\circ}$  when wearing the Hololens 1 compared to the No HMD condition ( $t = 2.54$ ,  $p = 0.03$ ). While the difference between the Quest 2 and the No HMD conditions was not statistically significant ( $t = 2.23$ ,  $p = 0.07$ ) for great circle angle error, the difference in adjusted mean values between the Quest 2 and Hololens 1 conditions was only  $0.1^{\circ}$ .

In the case of lateral angle error, participants demonstrated the worst performance in the No HMD condition, which was significantly  $1.2^{\circ}$  different from the Quest 2 condition ( $t = 2.53$ ,  $p = 0.03$ ). This result was unexpected as the participants were expected to exhibit the best performance in the No HMD condition.

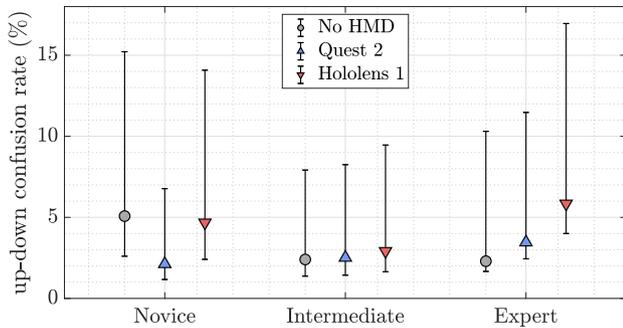
While the main effect of localization task experience was not significant for any of the angular metrics, the interaction between HMD condition and localization experience was significant for local great circle angle error ( $F = 4.40$ ,  $p = 0.001$ ), though it was not significant for lateral angle or weighted-polar angle errors. As seen in Figure 6, expert participants performed better in the the No HMD condition than either the Quest 2 ( $t = 3.45$ ,  $p = 0.02$ ) or the Hololens 1 ( $t = 3.81$ ,  $p = 0.004$ ). However, both the novice and intermediate participants had statistically similar performance across all three HMD conditions. The significance of this interaction suggests that while the HMDs may affect the localization perception of sound sources at familiar locations, the amount of error introduced by the HMDs does not exceed the error of localizing unfamiliar sound source locations.



**Figure 7.** Localization performance (text values) across speaker positions, depicting up-down confusion rates for the (a) Quest 2 and (b) Hololens 1 conditions. Metric values are reported above each speaker position (white circles), around which Voronoi cells have been drawn to simplify results interpretation.

### 3.3 Impact of HMD on localization reversal errors

The influence of HMD on front-back and up-down confusion rates was examined in two separate repeated measures logistic regression GLMMs that included factors of HMD condition and localization task experience. While the HMD condition had no significant impact on front-back confusions, it did significantly affect up-down confusion rates ( $\chi^2(2) = 10.37$ ,  $p = 0.006$ ). Pairwise comparison of Tukey-adjusted means indicated that the Hololens 1 induced a higher up-down confusion rate (2.4%) compared to the Quest 2 (1.3%) and the No HMD (1.5%) conditions (see Tab. 1). 2.4% amounted here to an average of less than 3 confused responses out of 96 tested positions per session. Up-down confusion rates for the Quest 2 and Hololens 1 conditions are illustrated for each of the 32 speaker positions in Figures 7a and 7b respectively.



**Figure 8.** Up-down confusion rate across HMD conditions and participants self-reported localization expertise.

There was also a significant interaction effect between HMD condition and localization task experience on the rate of up-down reversals ( $\chi^2(4) = 14.11, p = 0.007$ ), shown in Figure 8. In a Tukey-means comparison test of the interaction, the only significant difference between the up-down confusion rates was found for the novice participants between the No HMD and the Quest 2 conditions ( $t = 3.42, p = 0.01$ ) and the Hololens 1 and the Quest 2 conditions ( $t = 3.07, p = 0.04$ ). Although it looks like there was a slight trend that the expert participants performed better without an HMD than with the Hololens 1, the means were not statistically different ( $t = -2.29, p = 0.28$ ).

Based on these results, the Hololens 1 induced more up-down confusions for the participants than the Quest 2 or No HMD conditions. Pending further investigations, one may attribute this slight increase to the Hololens 1 headband protruding close to and over the ears, potentially distorting the incoming sound waves compared to the other two conditions. The percentage of up-down reversals was still quite small and may not drastically affect listener experience.

### 3.4 Procedural learning assessment

The impact of set number (i.e. first, second, or third set of the localization task) was evaluated to verify that a learning or fatigue effect did not occur over the course of the experiment. An improvement of participants performance during the experiment would suggest some procedural learning, as they were not provided with feedback during the task. A decrease in performance on the other hand would be attributed to a fatigue effect, caused by the duration and the difficulty of the task.

Repeated measures ANOVAs were used to assess whether set number significantly affected the angular error metrics introduced in Section 2.3. The results indicated that there were no significant effect of set number for any of the angular metrics ( $p > 0.05$ ). A similar analysis was conducted on front-back and up-down confusion rates, using repeated measures GLMMs [15] and likelihood ratio tests to assess the significance of the factor of set number. Once again, there was no effect of set number on either localization reversal rates ( $p > 0.05$ ). Since the participants did not tend to improve or worsen in localization

performance over the course of the experiment, the small amount of training at the beginning of the experiment seems to have been sufficient in providing procedural learning for the participants.

## 4 Summary and conclusion

In the present study, the impact of two HMD types on localization performance for 32 real sources distributed over the whole sphere was examined using several localization performance metrics. The results demonstrate that: (1) the HMDs significantly affected the local great circle angle error for participants that were familiar with the speaker positions, but otherwise did not introduce more error when the participants were unfamiliar with the source locations, and (2) the Hololens 1 increased the participants' up-down confusion rate by 0.9–1.1% compared to the other two conditions. Comparing these results to the observations in Ahrens et al. [3] that showed that the bulky HTC Vive 1 HMD significantly increased the elevation error by 1.8°, both bulky and glass-visor HMDs may affect the perception of sound source elevation. However, the extent of the effect is relatively small, as reported here and in Lladó et al. [9], and most likely will be masked by the difficulty of the localization task.

These results are valuable for future studies conducted in AR environments, especially when both real and virtual sources are intended to be localized while wearing an HMD. Wearing an HMD has a sufficiently small impact on real source localization that it can safely be considered as an HMD-free condition in most but the most demanding AR auditory localization studies.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgments

This study received approval from the *Comité d'Éthique de la Recherche*, Sorbonne University, project number CER-2021-094. This work was carried out in the context of the SONICOM project (<https://www.sonicom.eu/www.sonicom.eu>) that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101017743. We wish to thank Dr. Brian F.G. Katz who provided insight and expertise that greatly assisted the research and improved the manuscript.

## References

1. J. Blauert: *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, 1997.
2. S. Carlile, K. Balachandar, H. Kelly: Accommodating to new ears: the effects of sensory and sensory-motor feedback. *Journal of the Acoustical Society of America* 135, 4 (2014) 2002–2011. <https://doi.org/10.1121/1.4868369>.

3. A. Ahrens, K.D. Lund, M. Marschall, T. Dau: Sound source localization with varying amount of visual information in virtual reality. *PLoS One* 14, 3 (2019) 1–19. <https://doi.org/10.1371/journal.pone.0214603>.
4. M. Cuevas-Rodriguez, D.L. Alon, S. Clapp, P.W. Robinson, R. Mehra: Evaluation of the effect of head-mounted display on individualized head-related transfer functions, in *International Congress on Acoustics*, 2019, pp. 2635–2642.
5. A. Genovese, A. Roginska: HMDiR: an HRTF dataset measured on a mannequin wearing XR devices, in *Audio Engineering Society Conference*, Audio Engineering Society, 2019, pp. 2–4.
6. A. Genovese, G. Zalles, G. Reardon, A. Roginska: Acoustic perturbations in HRTFs measured on mixed reality headsets, in *Audio Engineering Society Conference*, vol. 8–4, Audio Engineering Society, 2018, pp. 2–16.
7. R. Gupta, R. Ranjan, J. He, G. Woon-Seng: Investigation of effect of VR/AR headgear on head related transfer functions for natural listening, in *Audio Engineering Society Conference*, Vol. 3–9, Audio Engineering Society, 2018, pp. 2–10.
8. C. Pörschmann, J.M. Arend, R. Gillioz: How wearing headgear affects measured head-related transfer functions, in *EAA Spatial Audio Signal Processing Symposium*, 2019, pp. 49–54. <https://doi.org/10.25836/sasp.2019.27>.
9. P. Lladó, T. McKenzie, N. Meyer-Kahlen, S.J. Schlecht: Predicting perceptual transparency of head-worn devices. *Journal of the Audio Engineering Society* 70, 7/8 (2022) 585–600.
10. G. Parsehian, B.F.G. Katz: Rapid head-related transfer function adaptation using a virtual auditory environment. *Journal of the Acoustical Society of America* 131, 4 (2012) 2948–2957. <https://doi.org/10.1121/1.3687448>.
11. D.R. Begault, E.M. Wenzel, M.R. Anderson: Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Acoustical Society of America* 490, 10 (2001) 904–916.
12. H. Bahu, T. Carpentier, M. Noisternig, O. Warusfel: Comparison of different egocentric pointing methods for 3D sound localization experiments. *Acta Acustica* 102, 1 (2016) 107–118. <https://doi.org/10.3813/AAA.918928>.
13. P. Leong, S. Carlile: Methods for spherical data analysis and visualization. *Journal of Neuroscience Methods* 80, 2 (1998) 191–200. [https://doi.org/10.1016/S0165-0270\(97\)00201-X](https://doi.org/10.1016/S0165-0270(97)00201-X).
14. P. Majdak, M.J. Goupell, B. Laback: 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Attention, Perception, & Psychophysics* 72, 2 (2010) 454–469. <https://doi.org/10.3758/APP.72.2.454>.
15. J.J. Faraway: *Extending Linear Models with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, Chapman & Hall, CRC, 2006.
16. R Core Team: *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. <https://www.R-project.org/>.
17. F. Zagala, M. Noisternig, B.F.G. Katz: Comparison of direct and indirect perceptual head-related transfer function selection methods. *Journal of the Acoustical Society of America* 147, 5 (2020) 3376–3389. <https://doi.org/10.1121/10.0001183>.
18. D. Poirier-Quinot, M.S. Lawless, P. Stitt, B.F.G. Katz: HRTF performance evaluation: Methodology and metrics for localisation accuracy and learning assessment. In: Dr. B.F.G. Katz, Dr. P. Majdak, Eds., *Advances in Fundamental and Applied Research on Spatial Audio*, Chapter 2, IntechOpen, Rijeka, 2022. <https://doi.org/10.5772/intechopen.104931>.
19. B. Rakerd, W.M. Hartmann: Localization of sound in rooms, II: The effects of a single reflecting surface, *Journal of the Acoustical Society of America* 78, 2 (1985) 524–533. <https://doi.org/10.1121/1.392474>.
20. P. Stitt, L. Picinali, B.F.G. Katz: Auditory accommodation to poorly matched non-individual spectral localization cues through active learning. *Scientific Reports* 9, 1 (2019) 1–14. <https://doi.org/10.1038/s41598-018-37873-0>.

**Cite this article as:** Poirier-Quinot D. & Lawless M. S. 2023. Impact of wearing a head-mounted display on localization accuracy of real sound sources. *Acta Acustica*, 7, 3.