Archives of Acoustics Vol. ***, No. ***, https://doi.org/10.24423/archacoust.2026.4326

Online First July 20, 2026

Research Paper

The Impact of the VR Interface on the Accuracy and Ergonomics of Spatial Hearing Tests

Grzegorz RUSINEK, Agnieszka Paula PIETRZAK*

Institute of Radioelectronics and Multimedia Technology, Warsaw University of Technology
Warsaw, Poland

*Corresponding Author: agnieszka.pietrzak@pw.edu.pl

Received September 8, 2025; revised March 6, 2026; accepted May 5, 2026; version of record July 20, 2026; published issue XXXX.

In recent years, virtual reality (VR) has been an attractive alternative to traditional methods of conducting hearing tests, offering both a natural way to indicate sound direction and complete digital data recording. The aim of this study was to analyze how the use of VR goggles and motion controllers affects the accuracy of localizing acoustic stimuli and the subjective comfort of participants, compared to a WEB-based application operated with a mouse. The study was performed for 16 azimuthal samples (0° elevation) and 8 elevation samples at azimuth 90°. Mean absolute angular errors (MAAE) were analyzed, and the paired-samples t-tests were performed. The results show that using the VR interface can reduce horizontal MAAE by 7° compared to the WEB-based interface. In the elevation tasks, no statistically significant differences were observed. Respondents rated VR the highest in terms of intuitiveness, comfort, speed, and perceived precision. These findings confirm the potential of VR to be used in localization tests, especially in the horizontal plane, and suggest directions for further research on HRTF personalization and interface optimization for elevation tasks.

Keywords: virtual reality (VR), spatial sound localization, user experience.

1. Introduction

Spatial audio quality listening tests often involve sound source localization. Participants hear short stimuli and then indicate the direction from which the sound originates. Although this simple concept stayed mostly unchanged throughout the years of research, researchers have tried multiple methods of gathering results from the participants.

For the localization task, experiments require an external response that maps participants' judgments onto measurable 3D coordinates. Many traditional methods are intuitive for listeners but require a supervisor to observe and transcribe responses, limiting scalability. Such methods may include: voice description of direction (Mason *et al.*, 2001), pointing with head/nose (Haber *et al.*, 1993; Bronkhorst, 1995; Fujii *et al.*, 2007), pointing with the torso/chest (Haber *et al.*, 1993; Fujii *et al.*, 2007), indicating direction with hand or with an auxiliary instrument, such as a baton (Haber *et al.*, 1993; Oldfield, Parker, 1984). These methods are comfortable from the perspective of the test participants, since their only task is to point in the direction of sound, which is an intuitive task to do. However, it is problematic in terms of collecting results, since in most cases it requires another person (a supervisor) actively observing the test participant, reading the indicated direction and recording the results.

Desktop and WEB-based tests often replace physical pointing with a mouse or gamepad that rotates a virtual sphere, providing broad reach but at the cost of indirect, cursor-driven mapping. Although convenient for large-scale data collection, such interfaces typically inflate completion times and, for elevation, report angular errors well above laboratory baselines.

Commodity head-mounted displays (HMDs) with six-degree-of-freedom (6-DoF) wands that track hand position with sub-millimeter precision allow natural reach-and-point actions. For the participant, pointing with the virtual reality (VR) controller is as intuitive as pointing with their hand. The controller can read its own rotation and position in space, which are then processed by a computer program, making it possible to determine the position of the sound source indicated by the test subject with sufficient precision. Ahrens *et al.* (2019) demonstrated that simply visualizing the controller tip as a laser beam inside the scene significantly reduced the elevation error compared with an HMD-only baseline. Valzolgher *et al.* (2023) examined the learning effects, showing that embodied actions (pointing or reaching) accelerated adaptation to spectral perturbations, yet they did not benchmark absolute localization performance against a browser-based task.

Previous research has contrasted different response modes within VR (laser versus reaching, natural versus symbolic actions) or compared VR to loudspeaker rigs without a corresponding graphical-user-interface (GUI) control. No study has systematically benchmarked a consumer VR pointing controller against a standard web/gamepad interface under matched acoustic rendering, using identical target sets and collecting both objective and subjective metrics. Such results can help the evidence-based decisions when choosing between low-cost browser tests and immersive setups for large-scale spatial-hearing assessments.

The aim of this study was to compare the VR pointing interface with a conventional mouse-driven WEB-based interface. Thirty-three non-expert listeners performed horizontal and vertical localization tasks using the same generic head-related transfer function (HRTF) stimuli, presented through the same headphones, with the difference only in the response modality. The main goal was to assess whether the use of the VR interface would result in the lower mean absolute angular errors (MAAE) than the WEB interface (in the horizontal or vertical plane) and whether participants would rate the VR interface higher in terms of intuitiveness, comfort, speed, and perceived precision.

2. Methodology

In this study, a listening test was designed and conducted to examine the effect of using the VR interface on the accuracy of virtual sound source localization, compared to a WEB-based application interface operated by a mouse. Additionally, a questionnaire survey was conducted to examine the subjective comfort of the participants when using the VR interface compared to the WEB-based one.

Thirty-three voluntary participants (10 female, 23 male) aged 18–26 years (M = 21.4, SD = 2.1), with no self-reported hearing problems completed the test. All of the participants were non-experts. Handedness was not assessed, but participants were asked to operate each interface with their dominant hand.

All participants took a virtual sound source localization test, using the VR interface and the WEB interface. The interface order was randomized. A break ranging from 5 min to 10 min was provided between blocks. The listening test consisted of twenty-four samples for each interface. Sixteen samples were presented in the horizontal plane (0° elevation) at fixed azimuth angles of 45°, 90°, 105°, 165°, 225°, 270°, 315°, 345°, each presented twice in random order. Eight samples were presented in the vertical plane (azimuth of 90°) at fixed elevation angles of −60°, −45°, 30°, 90°, each presented twice in random order. Among the twenty-four samples, half were white noise samples (eight samples with variable azimuth and four with variable elevation), and the other half were samples of female voice with analogous division into samples with variable azimuth and elevation. Two stimulus types (white noise and speech) were included to reduce task predictability. Each time, the samples were played in random order, but the samples with variable azimuth were played first, followed by those with variable elevation. For the test in both forms (using VR and using the WEB application), the same audio samples were used. The hardware used for the VR test were the Meta Quest 2 VR goggles with their dedicated spatial motion controllers with Audio-Technica ATH-M50X closed-back headphones. The results were compared with the same test conducted using a WEB application on a PC computer with the same Audio-Technica ATH-M50X closed-back headphones. Stimuli were presented at the sound level adjusted by the user, with inter-channel level matched within ±0.5 dB. Rendering used third-order ambisonics (fs = 48 kHz) and the IEM Binaural Decoder (default HRTF with headphone compensation for ATH-M50x).

In the case of the WEB application, the participants were asked to mark the sound sources on a 3D simulation displayed on a monitor, with an interactive image of a listening dummy and a sphere-shaped cage surrounding it, on which possible locations of the sound source were marked (Fig. 1). The response grid had 15° resolution in both azimuth and elevation. The participant had to click on one of the possible locations to give an answer. When conducting a test using VR, the test subject was placed in a virtual scene that consisted of a vast plain, with a virtual test stand placed in the middle. The stand contained a large sphere-shaped cage, with the participant placed in the middle of it (Fig. 2). The sphere-shaped cage helped with better orientation in space and corresponded to the cage used in the WEB application. Participants responded by pointing at one of the predefined locations on the grid (ray-casting) and confirming their choice by pressing a button on the virtual stand. Because the binaural scene was head-locked and did not update with head orientation, head tracking was not used in either condition, and participants were instructed to keep their heads still while listening.

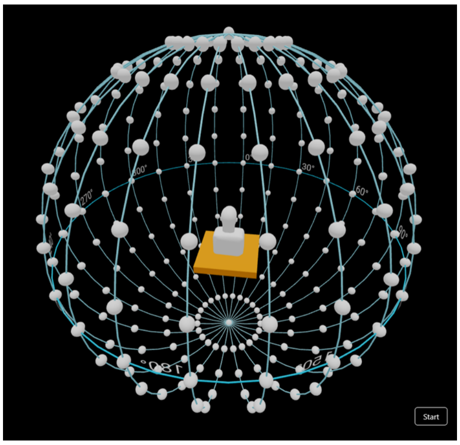

Fig. 1. View of the WEB sound-localization interface.

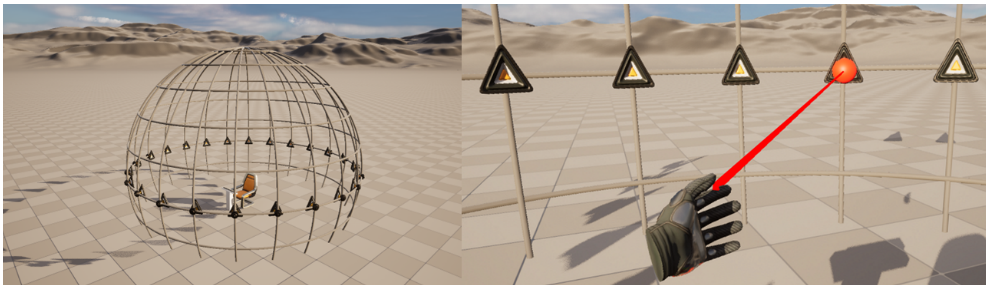

Fig. 2. View of the VR sound-localization interface.

The primary outcome was the MAAE [°], computed per participant separately for VR and WEB in the horizontal (H) and vertical (V) planes. The results were compared for VR and WEB within subjects using two-tailed paired t-tests for H, V, and the per-participant mean of $H + V$. Because each spatial position was presented only twice, the number of observations per angle was considered too small to provide reliable statistical comparisons at the level of individual spatial positions. Therefore, the statistical analysis focused on plane-level performance.

When conducting the test, special emphasis was also placed on comparing the performance of participants in the task with both approaches (does the use of VR with its innovative input method improve the ability of localizing sound over the conventional way of performing the test?) and also gathering feedback from participants regarding the methods of conducting the test – both in terms of their perceived performance in the task and their personal method of preference based on comfort, ease of use and other factors.

3. Results

Localization accuracy was assessed for 33 listeners who completed the task with both interfaces: immersive VR (head-mounted display + handheld controller) and conventional WEB (desktop display + mouse). Separate analyses were carried out for the horizontal (azimuth) and vertical (elevation) planes, followed by an overall $(H + V)$ comparison.

All participants' data were analysed for the identification of potentially 'unreliable' listeners (MAAE > mean + 2SD), and although partial results of two participants exceeded the considered threshold, a closer examination revealed no significant impact of this data on the overall conclusions. Furthermore, these participants did not report any issues during the tests. Therefore, especially because one of the goals was to assess the intuitiveness and ease of use of the two interfaces, it was ultimately decided to retain the data of all participants ($N = 33$) for further analysis.

3.1. Localization in horizontal plane

The distribution of horizontal-plane localization errors (MAAE) obtained with the two test interfaces is presented in Fig. 3. Results for the test performed with the use of the VR interface show a noticeably lower median ($M = 39.9°$) and a narrower distribution of errors than for the WEB interface, for which the median $M = 46.9°$ and the upper whisker reaches almost 90°. It suggests that the listeners not only made smaller azimuth errors with the immersive headset and hand-held controller, but they also did so more consistently, whereas the mouse-based WEB interface produced a broader spread that includes several exceptionally large errors.

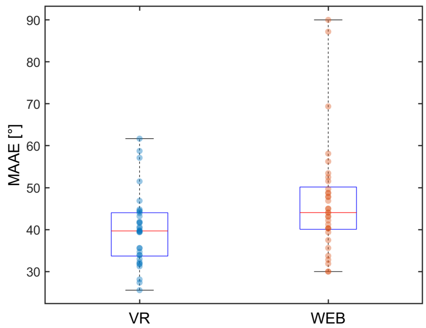

Fig. 3. Boxplots of MAAE [°] in the horizontal plane across all participants for the VR and WEB interfaces, with individual participant data points overlaid.

In the horizontal plane participants localized sounds by 7.0° more accurately in VR than on the WEB interface (Table 1). The paired-sample t-test (Table 2) revealed a statistically significant difference, ($t(32) = -2.65$, $p = 0.012$), which confirms a meaningful VR advantage in the azimuth task.

Table 1. MAAE and SD for the VR and WEB interfaces in the horizontal, vertical, and combined planes.

	MAAE [°]		SD [°]		Δ MAAE [°]
Plane	VR	WEB	VR	WEB	VR-WEB
Horizontal	39.87	46.90	8.59	13.68	−7.03
Vertical	45.44	40.97	10.62	13.66	4.47
$H + V$	42.66	43.93	5.59	11.15	−1.28

Table 2. Paired comparisons of MAAE (VR-WEB) across planes, mean difference (Δ), 95 % CI, paired t-test and Wilcoxon signed-rank statistics.

	Δ MAAE [°]	CI95_L	CI95_H	t-paired	p-value	Cohen's dz	W	p-value
Plane	VR-WEB							
Horizontal	−7.03	−12.44	−1.62	−2.65	0.012	−0.46	150.5	0.020
Vertical	4.47	−0.63	9.58	1.78	0.084	0.31	379	0.078
$H + V$	−1.28	−5.04	2.48	−0.69	0.494	−0.12	257	0.675

3.2. Localization in vertical plane

Figure 4 compares vertical-plane localization errors for the two considered interfaces. In this elevation task the pattern reverses: the WEB interface exhibits a slightly lower median MAAE ($M = 41.0°$) than the VR set-up ($M = 45.4°$), indicating that, on average, listeners judged elevation slightly more accurately with the mouse-based GUI. At the same time the WEB distribution is wider, signifying greater inter-individual variability. By contrast, the VR errors have a narrower inter-quartile range, suggesting more consistent, although less accurate, performance for the elevation angles.

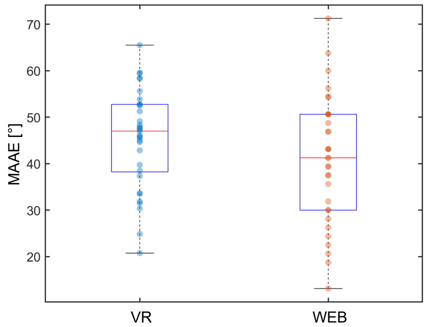

Fig. 4. Boxplots of MAAE [°] in the vertical plane across all participants for the VR and WEB interfaces, with individual participant data points overlaid.

In the vertical plane, the MAAE was by 4.5° lower for the WEB than the VR interface (Table 1). However, neither the parametric ($p = 0.084$) nor the non-parametric ($p = 0.078$) test (Table 2) reached the 0.05 significance threshold, suggesting that any vertical-plane advantage of the WEB interface is rather reflecting inter-individual variability rather than a systematic effect.

3.3. Overall localization comparison

When horizontal and vertical errors were averaged for each participant to obtain a single overall-MAAE score, the two interfaces converged. The VR interface resulted in the MAAE of 42.7° (SD = 5.6°), whereas the mouse-based WEB interface produced 43.9° (SD = 11.2°). Thus, the overall difference was −1.3°, far smaller than the typical within-listener spread. What could be more informative is the difference in the standard deviation and inter-quartile range: the distribution of errors for the VR interface is noticeably narrower. However, the overall analysis suggests that the significant horizontal advantage of VR (≈7°) is mostly offset by its moderate vertical shortfall (≈5°), leaving overall accuracy virtually unchanged yet noticeably more consistent in the immersive condition. This balance explains why statistical test failed to reach significance, and it underscores that the overall benefit of VR may lie in reducing inter-individual variability rather than in shifting the grand mean.

3.4. Perceived usability and participant experience

Immediately after completing the localization test with both interfaces, participants rated four aspects of perceived usability on 5-point Likert scales (1 = very poor, 5 = very good): intuitiveness, comfort, accuracy (perceived, meaning the confidence in one's indications), and speed. For each interface an overall usability score was also computed as the mean of these four items. Participants were also asked to choose their preference between VR and WEB interface, and leave an open comment if necessary.

The ratings show that participants favored VR over WEB across all aspects (Table 3). VR received higher ratings for intuitiveness (4.58 vs 3.94), comfort (4.45 vs 3.88), perceived accuracy (4.06 vs 3.48), and operating speed (4.09 vs 3.58). Paired t-tests ($N = 33$) indicated significantly higher ratings for VR on all aspects.

Based on per-participant averages, 73 % preferred VR, 18 % preferred the WEB interface, and 9 % reported no preference. In open comments participants most often praised the 'natural' aiming with the VR controller and cited 'difficulty achieving precise clicks' with the WEB cursor. Occasional remarks referred to the slight page

Table 3. Mean Likert-scale ratings (1–5) for usability aspects comparing VR and WEB; Δ = VR-WEB, with paired t-tests indicating significantly higher scores for VR on all items, including overall usability.

Aspect	VR	WEB	Δ (VR-WEB)	t-paired	p-value
Intuitiveness	4.58	3.94	0.64	3.37	0.002
Comfort	4.45	3.88	0.58	3.03	0.005
Accuracy (perceived)	4.06	3.48	0.58	2.70	0.011
Speed	4.09	3.58	0.52	2.52	0.017
Overall usability score	4.30	3.72	0.576	4.62	$<.001$

latency or headset pressure. These findings support VR as the preferred interaction modality for spatial hearing experiments, particularly when naturalistic pointing and efficient supervision are a priority.

4. Discussion

Two interaction modalities used in spatial sound localization experiments were compared: a mouse-driven WEB interface and an immersive VR interface using a handheld controller. While both conditions relied on identical binaural stimuli and headphones, the response method differed. The results demonstrate that the VR interface improved localization accuracy in the horizontal plane, whereas no statistically significant advantage was observed in the vertical plane.

The improvement in horizontal localization accuracy when using the VR interface may be explained primarily by the more natural response modality. In the immersive condition, participants indicated the perceived sound direction by pointing with a tracked controller, which closely resembles natural reaching and pointing actions. Previous studies have shown that embodied interaction can facilitate spatial perception tasks and reduce cognitive load associated with translating auditory perception into a response (Valzolgher *et al.*, 2023). In contrast, the The WEB interface required participants to translate a perceived direction into cursor movement on a 2D display, which introduces an additional visuomotor mapping step. Such indirect interaction may increase response uncertainty and contribute to larger angular errors.

This advantage of the VR interface was observed only for the horizontal plane. In the elevation task, the WEB interface yielded slightly lower mean errors, although the difference was not statistically significant. One possible explanation is that vertical localization relies strongly on spectral cues encoded in the head-related transfer function (HRTF). Because the present experiment used a generic HRTF, listeners may have experienced front-back or elevation ambiguities that cannot be easily compensated by improved interaction methods. Previous research has shown that elevation localization remains challenging when using non-individualized HRTFs, even in immersive environments. Another relevant factor is the absence of dynamic auditory cues. In the present implementation the binaural scene was head-locked and did not update with head movements. Head motion is known to provide additional information that helps disambiguate elevation and front-back confusions. Studies in spatial hearing have shown that even small head movements can significantly improve localization performance by providing dynamic changes in interaural cues and spectral filtering. Because head tracking was intentionally disabled in this experiment to maintain identical acoustic rendering across conditions, participants could not benefit from these dynamic cues. This limitation may have particularly affected elevation judgments.

Despite the limited improvement in objective accuracy, the subjective evaluations clearly favored the VR interface. Participants rated VR higher in terms of intuitiveness, comfort, perceived accuracy, and operating speed, and the majority expressed a preference for the immersive condition. These findings suggest that response method that resembles natural pointing may reduce task difficulty and increase participant engagement, which can be beneficial for large-scale listening studies. The results also suggest that the VR interfaces may produce a narrower distribution of errors among participants, indicating more consistent performance.

Several limitations of this study should be acknowledged. First, the use of generic HRTFs may have limited elevation accuracy and introduced perceptual ambiguities. Second, the binaural rendering was head-locked, preventing the use of dynamic auditory cues that are known to improve spatial perception. Third, the participant sample consisted mainly of young non-expert listeners, which may limit the generalizability of the results. Finally, the number of tested spatial positions was relatively small and constrained to discrete grid points. Future work should investigate localization performance in VR environments that include head-tracked dynamic binaural rendering and individualized HRTFs where possible. Such improvements could allow participants to utilize natural head movements and more accurate spectral cues, potentially leading to improved elevation localization. Additionally, larger and more diverse participant groups would help determine whether the advantages of the VR interfaces generalize across populations and experimental contexts.

5. Conclusions

This study compared an immersive virtual-reality interface with a conventional mouse-driven WEB application for spatial sound-localization tasks. The VR condition reduced the horizontal mean-absolute error by 7° (≈15 %). No statistically significant advantage was found in the vertical plane, suggesting that elevation cues still pose a challenge when rendered with a generic HRTF. Subjective ratings favored VR on intuitiveness, comfort, speed and perceived accuracy, with 73 % of participants preferring the immersive setup.

While the findings confirm that natural pointing with motion controllers can improve the ergonomics of psychoacoustic testing, the overall localization accuracy remained limited and the vertical performance did not improve. Future work should examine the use of individualised HRTFs where possible and include larger, more diverse samples to validate the observed horizontal-plane benefit. Because in this study the VR implementation was head-locked, which is also a limitation, future works could enable world-stabilized, head-tracked micromovements to quantify the expected gain in elevation accuracy and test whether VR's advantage grows when dynamic cues are available. An optimised VR protocol could provide researchers with a scalable, supervisor-independent platform for high-quality spatial-hearing experiments.

Fundings

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contributions

Agnieszka Paula Pietrzak conceptualized the study, supervised the project, and wrote the original draft. Grzegorz Rusinek developed the software (WEB/VR implementation). Both authors designed the methodology and the experiment, curated the data, performed the formal analysis, and prepared the figures and tables, reviewed and edited the manuscript. All authors have accepted the responsibility for the entire content of this manuscript and approved its submission.

References

1. Ahrens A., Lund K.D., Marschall M., Dau T. (2019), Sound source localization with varying amount of visual information in virtual reality, *PLOS ONE*, **14**(3): e0214603, https://doi.org/10.1371/journal.pone.0214603.

2. Bronkhorst A.W. (1995), Localization of real and virtual sound sources, *The Journal of the Acoustical Society of America*, **98**(5): 2542–2553, https://doi.org/10.1121/1.413219.

3. Fujii A., Ohsugi Y., Yamamoto Y., Nakamura T., Sugiura T., Tauchi M. (2007), Comparison of different pointing methods for sound localizability measurement in the vision impaired subjects, *Journal of Physiological Anthropology*, **26**(3): 381–385, https://doi.org/10.2114/jpa2.26.381.

4. Haber L., Haber R.N., Penningroth S., Novak K., Radgowski H. (1993), Comparison of nine methods of indicating the direction to objects: data from blind adults, *Perception*, **22**(1): 35–47, https://doi.org/10.1068/p220035.

5. Mason R., Ford N., Rumsey F., de Bruyn B. (2001), Verbal and nonverbal elicitation techniques in the subjective assessment of spatial sound reproduction, *Journal of the Audio Engineering Society*, **49**(5): 366–384.

6. Oldfield S.R., Parker S.P.A. (1984), Acuity of sound localisation: a topography of auditory space. I. Normal hearing conditions, *Perception*, **13**(5): 581–600, https://doi.org/10.1068/p130581.

7. Valzolgher C., Capra S., Sum K., Pavani F., Picinali L. (2023), Training spatial hearing skills in virtual reality through a sound-reaching task, [in:] *Forum Acusticum 2023* (Torino, Italy), https://doi.org/10.61782/fa.2023.0341.