The Effects of Display Devices on Locating Sound Sources in Virtual Environments

Tibor Guzsvinecz

Department of Information Technology and its Applications, Faculty of Information Technology, University of Pannonia,
Gasparich M. utca 18/A, 8900 Zalaegerszeg, Hungary
Hungarian Research Network, Piarista u. 4, 1052 Budapest, Hungary
guzsvinecz.tibor@zek.uni-pannon.hu

Abstract: Sound localization plays a crucial role in both real and virtual environments, particularly for effective navigation. Naturally, this ability can be affected by various factors. To understand how it is influenced, the sound localization skills of 110 students were measured within a virtual environment. Two versions of this virtual space were created, and all students used both. The desktop display version was used first, then the Gear VR version was used afterward. A Sony PlayStationTM Platinum Wireless headset was used on both versions. The results were grouped and examined by various human characteristics and display devices. According to the results, the following conclusions can be made: gender influenced sound localization, since males outperformed females using the desktop display, but not with the Gear VR. This was due to significant accuracy improvements for females with the mentioned head-mounted display. Previous virtual reality experience improved accuracy and speed with the Gear VR. The proportions of incorrect localizations differed significantly between the two versions, indicating that the immersion level affects them. These results can affect the design and implementation processes of future virtual reality experiences, in several fields of research and entertainment.

Keywords: cognitive skills; human-computer interaction; sound; virtual environment; virtual reality

1 Introduction

Both sound localization and visuospatial skills are essential for us to navigate safely and understand our surroundings in the real world. Although human auditory perception is generally less acute than visual perception [1], we can still use it to detect hazards, alarms, approaching vehicles, and so on [2]. Also, visual and auditory skills can enhance safety and risk communication [3]. Overall, localizing sounds is quite accurate under normal conditions [4]. However, it should be noted that distances over about 1.5 meters are usually underestimated, while those that are below it are usually overestimated [5] [6].

Similarly, the skill of sound localization is essential in virtual reality (VR) as well due to the fact that virtual environments (VEs) simulate real ones [7] [8]. Not to mention, we, as humans, are integral components of VR systems [9]. Cognition is also an important aspect of these systems [10] [11] since they can enhance our cognitive functions [12]. Thus, the target groups have to be identified, and focus should be put on them when developing such virtual spaces [13]. These VEs can be created for several purposes, such as healthcare [14], educational [15] and entertainment purposes [16].

As the design of VEs is a user-centered process, their composition, display devices, and human characteristics can affect performance and immersion [17, 18]. The case is similar regarding sound design [19]. Several studies in the literature show that there is a connection between visual and auditory cues [20-23]. It is also shown that VR-based sound localization training could yield positive results [24]. Yong and Wang [25], along with Cullen et al. [26], concluded that the presence and perception of sounds significantly enhance orientation in VEs. Another research yielded the following preliminary results [27]: 1) Correct sound sources were significantly more frequently located by males than females by 40.5%; 2) Females using the Gear VR exhibited 20.83% greater sound localization accuracy compared to those who used a desktop display; 3) Males using the Gear VR located sound sources significantly faster by 27.89% than those who used the desktop display. Other studies have shown that spatial sound contributes to immersion and user performance in virtual tasks [28] [29]. The process of sound localization depends on interaural time differences (ITDs), interaural level differences (ILDs), and head-related transfer functions (HRTFs), all of which may be affected by the design of VR devices. Moraes et al. evaluated physiological responses to auditory stimuli in VR and found that spatialized 3D audio improved localization accuracy and reduced cognitive strain, suggesting that audio fidelity is crucial in immersive systems [30]. Cognitive load can also influence sound localization performance. VR interfaces demand constant multisensory integration and may lead to sensory overload [31]. Wang et al. specifically demonstrated that auditory-assisted visual search in mixed reality environments increases task success but also imposes higher cognitive load, depending on the sound design and user training [29]. Reissmüller emphasizes that the integration of haptics and spatial audio in VR systems leads to a stronger sense of embodiment, but may increase the mental demands placed on users, particularly novices [32].

Therefore, understanding how sounds are localized in VEs is not an easy task. To gain a deeper understanding, a virtual space was created that can be used with a desktop display and the Gear VR. This setup allowed participants to experience the same VE under two distinct levels of immersion. The creation of this VE is made easier by the Cognitive InfoCommunications (CogInfoCom) environment since it puts focus on human-computer interaction [33-35]. For instance, Sudár and Horváth offered design recommendations for digital workflow-oriented desktop VR spaces [36]; Kővári studied the relationship between the Internet of Digital Education and VR [37]; and Guzsvinecz and Szűcs analyzed textual reviews of VR games to understand how users felt about them [38].

Overall, the goal of this research is to provide recommendations for researchers and developers of VR applications so that they can create a tailored experience for different people with different display devices. Thus, the following three research questions (RQs) are formed:

RQ1: Does immersion level affect the number of correct sound localization and its speed?

RQ2: Do human characteristics affect the number of correct sound localizations and their speed?

RQ3: Is the proportion of incorrect directions similar on each display device?

As can be seen, RQ1 addresses a core perceptual concern in VR usability. Since prior studies have shown that spatial hearing can be altered by immersive settings due to altered head movements, visual occlusion, and changes in egocentric reference frames. However, few empirical investigations have directly compared localization performance between flat-screen and immersive VR settings under controlled conditions. By isolating the device variable while keeping other parameters constant, this study can provide a deeper understanding of how device-mediated immersion may affect sound localization. Regarding RQ2, it is known that human characteristics and prior experience are known to influence spatial cognitive abilities in technologically mediated environments. Exploring these factors helps in identifying subgroups for whom specific device types may be more or less effective, thus advancing the goal of user-centered design in VR systems. Lastly, RQ3 frames the results in terms of practical applications. If Gear VR users, for instance, consistently underperform relative to desktop users in localizing certain sound angles, developers may need to compensate through enhanced spatial audio rendering or multisensory cue integration. Conversely, if immersive systems yield superior accuracy, these results can inform the optimization of auditory feedback in training simulations, serious games, or accessibility tools for the visually impaired.

The following step was to create three null hypotheses (Hs) that correspond to the previous three RQs. These three Hs are the following:

H1: Immersion level does not affect the number of correct sound localization and its speed.

H2: Human characteristics do not affect the number of correct sound localizations and their speed.

H3: The proportion of incorrect directions is similar between the two display devices.

The structure of the paper is the following. Section 2 presents the used materials and methods. Then, the results are detailed in Section 3. Afterwards, they are discussed in Section 4. The last step is to draw conclusions in Section 5.

2 Materials and Methods

This section contains three subsections. Subsection 2.1 details the development process of the VE. Data collection and analysis are shown in Subsections 2.2 and 2.3, respectively.

2.1 The Virtual Environment

The VE was developed using the Unity game engine (version 2018.4.22f1). Two versions were created of this environment: one for PC and one for Android. The former version requires a keyboard and mouse, while the latter needs the Gear VR to be used. Naturally, the PC version used a desktop display. Except for interaction and immersion level, the two versions are identical. To accurately replicate sounds within the virtual environment, two types of sounds were considered. These two were direct sounds and indirect (reflected) sounds. Naturally, the audio is controlled with the virtual head rotation of the user, meaning it changes its perceived location based on where the user is looking. This has to be recalibrated whenever the virtual camera shifts its orientation. Therefore, the volume of the sound also changes. A low-pass filter is also used, and it is affected by the position of the sound source.

It is vital to distinguish between sounds that are in front of and behind the user, since the human ear has the ability to block certain sound volumes. Notably, this is done in case of higher sound frequencies when the source is positioned behind the individual. The application splits the virtual space into four sections: front, left, right, and rear of the virtual camera. This enables a clearer perception of sounds that are typically dimmed by the user's head and ear. Sound becomes quieter if the user moves farther away from or turns away from the source.

In Figure 1, red lines are observable. These red lines represent points of sound reflection. The center of the virtual camera, where these red lines converge, corresponds to its position. This is the point where the location of the user can also be found within the virtual environment. It is important to note that these red lines serve an illustrative purpose only and are not visible to the user inside the VR space.

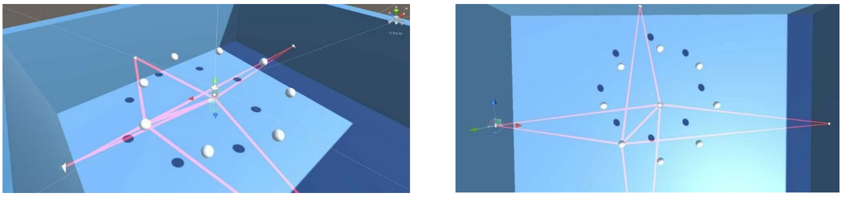


Figure 1
Two screenshots from the virtual environment, illustrating the reflection points

Upon launching the application, the user is placed in the main menu. Before starting any tests, users are required to provide certain information in the menu. These details are age, gender, dominant hand, glasses usage, and any past or present hearing difficulties. For those who used the Gear VR version, the following piece of information was needed: whether they have previous experience with virtual reality. Once all information is supplied, the testing phase can be started. When clicking on the start button, participants are transported into the VE.

2.2 Data Collection

In the VE, the users were greeted with the instructions on a panel. Subsequently, pressing the "Start" button started the test. There was a reticle at the center of the screen that functioned as a cursor. It turned red when interacting with clickable elements in the environment. As could be observed in Figure 1, eight spheres could be found around the virtual camera within the VE. These were 8.3 "Unity units" away from the camera, which represent 8.3 meters. These spheres are the potential sound sources. As the test began, a randomly chosen sphere started emitting sounds after a few seconds. This was where the measurement of the reaction times was started. The sound continued to play from the same source until one sphere was clicked. Also, the reaction time measurement for that subtest concluded with the click, and the accuracy of the response was displayed on the screen. Afterward, the room's dimensions randomly change between 100% and 250% to change acoustics. There were ten subtests, and after each, the following are logged into a file:

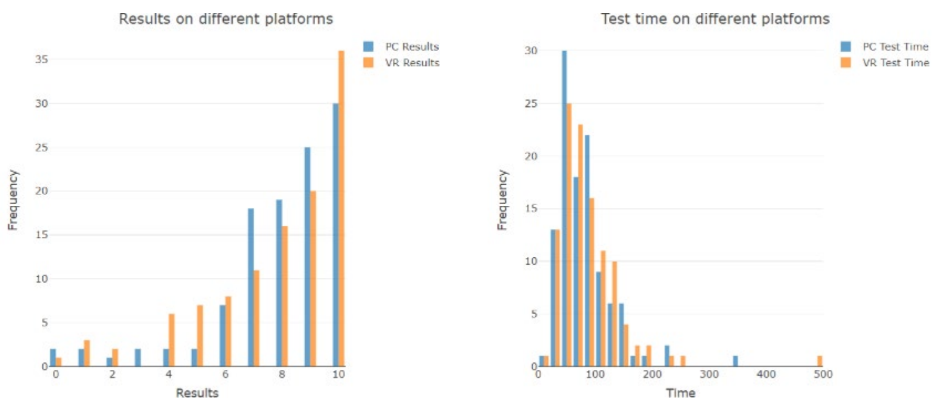
- Whether the response is correct or not
- If the response is incorrect, the relative position of the chosen answer in relation to the correct one
- The relative position of the correct answer compared to the previous correct response
- Room size
- Reaction time

The measurement process followed the following structure. Initially, it was conducted on a PC by one participant. This process usually took less than ten minutes. Then, the participant was given a break. Then, the measurement process proceeded with the Gear VR version, which also took around ten minutes. It is important to note that the sound sources within the virtual environment were randomized. This eliminated the possible issue of memorizing correct answers. Naturally, the participants could have questions during the measurement process. To address this, a researcher was constantly present in the room alongside the participants. Any questions regarding the tests or devices could be answered easily. During the tests, the participants used a “Sony PlayStation™ 4 Platinum Wireless” headset on both platforms. This headset was chosen because it could simulate virtual 7.1 sound.

The measurements started in the fall of 2019. The measurements quickly slowed down due to the COVID-19 pandemic. Thus, besides students, it started to involve friends, friends of friends, and colleagues. To ensure safety, each device was disinfected after each participant. The measurements concluded in the summer of 2022 with 110 participants. Out of 110 participants, 24 were female, and 86 were male. Their average age was 24.92, ranging from 11 to 61. The tests included 103 right-handed and 7 left-handed participants. 18 of the people had previous VR experience, while 59 participants wore glasses.

2.3 Data Analysis

After data collection, the R programming language was used to analyze the data. 0.05 was chosen as the alpha level for the data analysis. Before conducting statistical tests, the dataset underwent a preprocessing phase to ensure consistency and reliability. Raw responses were screened for completeness. Then, the histograms of results and test times on different platforms were calculated. These can be seen in Figure 2.



Histograms of results on PC and VR (left); data distribution of test time on PC and VR (right)

After assessing the frequency of the results and time on PC and VR, their distributions were examined with the Shapiro-Wilk normality test. The results show that the distributions of the results and time were not Gaussian:

- $W = 0.80551$, $p < 0.001$, in the case of results on PC
- $W = 0.83434$, $p < 0.001$, in the case of results on VR
- $W = 0.83030$, $p < 0.001$, in the case of test times on PC
- $W = 0.71824$, $p < 0.001$, in the case of test times on VR

However, since certain groups were assessed in this study, their distributions were also assessed. The results of the examination can be seen in Table 1.

Table 1

Results of the Shapiro-Wilk normality test for each group

	PC (answers)	**VR (answers)**	**PC (time)**	**VR (time)**
Male	W= 0.785, p < 0.001	W = 0.831, p < 0.001	W = 0.892, p < 0.001	W = 0.870, p < 0.001
Female	W= 0.909, p = 0.034	W = 0.848, p = 0.002	W = 0.758, p < 0.001	W = 0.604, p < 0.001
Right-handed	W= 0.812, p < 0.001	W = 0.844, p < 0.001	W = 0.821, p < 0.001	W = 0.712, p < 0.001
Left-handed	W= 0.759, p = 0.015	W = 0.738, p < 0.001	W = 0.828, p = 0.077	W = 0.644, p < 0.001
Glasses	W= 0.802, p < 0.001	W = 0.887, p < 0.001	W = 0.970, p = 0.158	W = 0.911, p < 0.001
No glasses	W= 0.798, p < 0.001	W = 0.747, p < 0.001	W = 0.805, p < 0.001	W = 0.587, p < 0.001
Hearing problems	W = 0.944, p = 0.683	W = 0.860, p = 0.261	W = 0.748, p = 0.037	W = 0.798, p = 0.098
No hearing problems	W = 0.804, p < 0.001	W = 0.832, p < 0.001	W = 0.804, p < 0.001	W = 0.683, p < 0.001
Previous VR experience	-	W = 0.812, p < 0.001	-	W = 0.939, p = 0.051
No previous VR experience		W = 0.858, p < 0.001		W = 0.725, p < 0.001

After addressing the distributions of the datasets, the following statistical tests were used: the Wilcoxon signed rank test when comparing the results between the two versions (as the participants were identical), and depending on the distribution, either the Wilcoxon rank sum test or the paired t-test was used when comparing the results between two groups of participants on the same platform. Inaccurate sound localizations were also assessed using the chi-squared test for given probabilities as well as Pearson's chi-squared test. The latter was used to understand how their proportions differ between the two types of immersion levels and to see whether the display devices themselves had significant effects on accuracy.

3 Results

This section is split into two subsections. In Subsection 3.1, sound localization accuracy is compared between the two versions. Then, in Subsection 3.2, the test times are compared between the two versions.

3.1 Comparing Sound Localization Accuracy

Before starting the examination, the general results should be examined. The dataset was divided into two sections based on the platform: PC and VR. Table 2 shows the descriptive statistics of the number of correct sound localizations.

Table 2

The descriptive statistics of accurately localizing sounds on PC and VR

	PC				VR			
	Min	Max	M	SD	Min	Max	M	SD
Male	1	10	8.41	1.78	0	10	7.82	2.44
Female	0	10	6.20	2.87	1	10	7.75	2.43
Right-handed	0	10	7.95	2.18	0	10	7.85	2.31
Left-handed	1	10	7.71	3.25	1	10	7.14	4.01
Glasses	0	10	7.83	2.32	0	10	7.32	2.55
No glasses	1	10	8.05	2.18	1	10	8.37	2.17
Hearing problems	7	9	8.00	0.81	5	10	8.00	2.44
No hearing problems	0	10	7.93	2.28	0	10	7.80	2.44
Previous VR experience	-	-	-	-	4	10	8.58	1.67
No previous VR experience	-	-	-	-	0	10	7.43	2.65

As can be seen in Table 2, the data suggest that certain human characteristics, such as gender and vision, may affect sound localization. However, using immersive display devices may reduce some differences. Additionally, having previous VR experience is a clear advantage. To understand whether there were significant differences between these groups, the aforementioned statistical tests were used.

Regarding gender, males generally outperformed females on PC (with the desktop display). The difference was significant, $W = 1562.5$, $p < 0.001$. However, this gap decreased in VR, since the sound localization accuracy of females was increased by 25% when they used the Gear VR. This increase was significant, $V = 33$, $p = 0.002$. When this gap decreased, the difference between the sound localization accuracy of males and females became not significant, $W = 1064.5$, $p = 0.812$. However, it also should be noted, that compared to the desktop display, the Gear VR significantly decreased the sound localization accuracy of males by 7.01%, $V = 1250$, $p = 0.012$.

Right-handed participants had similar accuracy on PC and VR. There were no significant differences between the two groups using either version (PC: W = 336, p = 0.764; VR: W = 337.5, p = 0.777). While the left-handed participants performed slightly worse in immersive environments compared to non-immersive ones, the difference in their sound localization accuracy was not significant, V = 6.5, p = 0.712. Similarly, right-handed participants had no significant difference between the immersion levels as well, V = 1635, p = 0.632.

Participants with glasses scored slightly lower on PC compared to those without glasses. This difference was not significant, W = 1400, p = 0.542. However, with the Gear VR, this difference was larger: as previously, those without glasses performed better than their counterparts who wore glasses. This difference, however, was significant between the two groups, W = 1110.5, p = 0.015. It was also assessed whether the use of the Gear VR significantly increased the sound localization accuracy of each group. The results of the examination show that the accuracy of neither group increased significantly (Glasses: V = 629, p = 0.112; No glasses: V = 315.5, p = 0.420).

Participants with hearing problems had high accuracy on both platforms. However, there was no significant difference between their accuracy when the platforms were compared, t(3) = 0, p = 1. Those without hearing problems also had similar accuracy using the desktop display and the Gear VR. They showed no significant difference between the two immersion levels, V = 1658.5, p = 0.551. Similarly, there were no significant differences between the accuracy of the two groups of participants on either display device (PC: W = 177.5, p = 0.579; VR: W = 224, p = 0.850).

Previous VR experience had a large impact on sound localization in immersive VR. Participants with previous VR experience significantly outperformed those who did not have such experience by 15.47%, W = 1655.5, p = 0.035.

Next, the proportions of inaccurate sound localizations were assessed on the PC version. This can be seen in Table 3.

Table 3

The number of inaccurate sound localizations on the PC platform

PC	**Behind**	**Behind, to the left**	**Behind, to the right**	**In front, to the left**	**In front, to the right**	**To the left**	**To the right**
Incorrect	24	8	6	77	92	11	9

As shown in Table 3, the number of inaccurate sound localizations was the largest in the case of the “in front, to the right”. Meanwhile, the second largest was in the case of the “in front, to the left”. This indicates that the errors most often occurred in neighboring spheres adjacent to the sound-emitting source. Thus, the two largest incidences of inaccurate sound localization, occurred in the case of the neighboring spheres. The next step was to see whether significant differences existed between their proportions. According to the results, such a difference existed, $\chi^2(6, N = 227) = 243.91, p < 0.001$.

The following step was to assess the number of inaccurate sound localizations on the VR platform. These numbers can be seen in Table 4.

Table 4
The number of inaccurate sound localizations on the VR platform

VR	Behind	Behind, to the left	Behind, to the right	In front, to the left	In front, to the right	To the left	To the right
Incorrect	8	9	8	114	94	3	5

As can be observed in Table 4, the same two directions yielded the two largest numbers of inaccurate sound localizations. However, contrary to the PC version, the largest number in this version was in the case of the "in front, to the left" and the second largest was in the case of the "in front, to the right". Next, their proportions were examined. Similarly, it yielded a significant difference, $\chi^2(6, N = 241) = 400.18, p < 0.001$.

The final step was to see whether the proportions of inaccurate sound localization differed between the two versions. With Pearson's chi-squared test, the following can be concluded: the proportions were significantly different; therefore, the platforms had a significant effect on the number of inaccurate sound localizations, $\chi^2(6, N = 468) = 20.848, p = 0.001$.

3.2 Comparing Sound Localization Times

The following step was to examine the sound localization times. First, the descriptive statistics were analyzed. They are shown in Table 5.

Table 5
The descriptive statistics of the localization times on PC and VR. The times are in seconds

	PC				VR			
	Min	**Max**	**M**	**SD**	**Min**	**Max**	**M**	**SD**
Male	19.2	229.8	74.7	40.3	17.4	242.1	78.5	43.1
Female	32.9	343.2	106.6	65.5	45.2	485.3	112.7	88.1
Right-handed	19.2	343.2	81.6	48.6	17.4	485.3	86.0	56.7
Left-handed	31.8	154.9	82.8	49.1	42.2	242.1	86.3	70.2
Glasses	19.2	154.9	75.5	32.0	17.4	242.1	84.9	46.2
No glasses	28.5	343.2	88.8	61.9	28.0	485.3	87.2	68.4
Hearing problems	19.2	154.9	114.2	64.0	17.4	242.1	136.2	120.1
No hearing problems	26.4	343.2	80.4	47.7	28.0	485.3	84.1	53.8
Previous VR experience	-	-	-	-	28.5	146.9	69.1	29.2
No previous VR experience	-	-	-	-	17.4	485.3	94.2	65.5

Males localized sounds faster than females on both platforms. On PC, this difference was significant, $W = 636, p = 0.004$. In the case of the VR version, this difference was also significant, W = 672, p = 0.009. However, neither the sound localization times of males (V = 1620, p = 0.281) nor females (V = 121, p = 0.422) were significantly different between the two versions.

Right-handed participants were slightly faster in sound localization than left-handed participants. This difference in localization times was neither significant on PC (W = 363, p = 0.980) nor on VR (W = 404, p = 0.598). Between the two versions, the two groups of users also had similar sound localization times (Right-handed participants: V = 2310, p = 0.226; Left-handed participants: V = 9, p = 0.468). Overall, these similarities between the two versions suggest a minimal impact of handedness on performance.

Participants wearing glasses localized sounds faster than those without glasses on PC, but the difference in time was not significant (W = 1480, p = 0.885). In VR, the difference became smaller between the two groups (W = 1541, p = 0.829). When comparing the same groups between the versions, the following conclusions could be made: those who did not wear glasses performed similarly between the two versions (V = 666, p = 0.981), and those who wore glasses had a weak significant difference in the context of localization times between the two versions (V = 625, p = 0.050).

Participants with hearing problems had longer localization times on both platforms. Still, according to the results of the Wilcoxon rank sum test, no significant differences were found between the two who had hearing problems and those who did not on both versions (PC: W = 294, p = 0.193; VR: W = 231, p = 0.767). The sound localization times of neither group of participants were significantly different between the two platforms (Those with hearing problems: V = 4, p = 0.875; Those without hearing problems: V = 2441, p = 0.214). This may be due to the small sample size.

In VR, participants with previous experience localized sounds significantly faster than those without such experience, W = 997, p = 0.033. This highlights the role of familiarity with VR technology in improving response times.

4 Discussion

This section is split into three subsections. Out of the three, the two subsections exist because H1 and H2 present mixed cases, and H3 is rejected. H1 and H2 are mixed cases because in some cases, correct sound localization and its speed are affected, and in some cases, they are not. H3 is rejected since the proportions of incorrect directions were different between the two platforms. Therefore, Subsection 4.1 presents the mixed cases, and the rejected hypothesis is detailed in Subsection 4.2. Lastly, the limitations of the study are shown in Subsection 4.3.

4.1 Mixed Cases

As mentioned, both H1 and H2 present mixed cases. Regarding H1, the following conclusions can be made. Firstly, gender differences were larger when a desktop display was used than with the Gear VR. Still, in both cases, males outperformed females in sound localization accuracy. However, the Gear VR improved the accuracy of female participants and thus narrowed the gender gap between males and females. Specifically, females exhibited a 25% increase in sound localization accuracy with the Gear VR. This suggests that immersive environments may provide cognitive or perceptual advantages that mitigate traditional gender-based differences in spatial abilities. Conversely, male participants had a slight decrease in accuracy with the Gear VR. These results align with prior research: when spatial abilities were assessed, the Gear VR similarly decreased the gap between the results of males and females [18]. Handedness showed minimal effects on sound localization accuracy: right-handed and left-handed participants showed similar results across both platforms. Those participants who wore glasses performed similarly to those without glasses on PC. However, their performance diverged significantly with the Gear VR since those with glasses had worse accuracy. This suggests that VR systems might amplify visual distractions or discomfort caused by wearing glasses. Participants with hearing problems performed similarly to those without hearing issues across both platforms. This indicates that sound localization tasks in this study were not significantly affected by mild hearing impairments. However, the small sample size of participants with hearing issues limits the generalizability of this conclusion. Lastly, participants with previous VR experience performed significantly better in terms of sound localization accuracy. This shows the importance of familiarity with immersive technologies in enhancing performance. These results emphasize the importance of training and user experience in VEs, mainly for tasks that rely on spatial cognition.

Regarding H2, the following can be concluded. Sound localization times were generally longer with the Gear VR compared to desktop displays. This may reflect the increased cognitive load associated with immersive environments. However, previous VR experience significantly reduced sound localization times. This, in turn, shows the importance of familiarity with such environments. These results suggest that while head-mounted displays provide enhanced immersion, they may also require more cognitive effort. Interestingly, the sound localization times of males were faster than those of females on both display devices. This suggests that gender differences in sound localization speed are robust to changes in display devices. However, the gap decreased when the Gear VR was used. This can emphasize the head-mounted display's potential to somewhat balance differences, contrary to non-immersive display devices.

4.2 Rejected Hypothesis

H3 was rejected since immersion levels had an effect on the proportions of inaccurate sound localizations. The distribution of such localizations significantly varied between the use of the desktop display on PC and the Gear VR in the VR versions.

Using the desktop display on the PC version, the largest proportion of errors occurred in the “in front, to the right” direction, followed by “in front, to the left.” This pattern suggests a systematic difficulty in accurately localizing sounds from the front plane, mainly near the right side of the auditory field. In contrast, the use of the Gear VR head-mounted display showed a reversal: “in front, to the left” produced the highest number of inaccurate sound localizations, followed by “in front, to the right.” This difference indicates that the immersive nature of VR influences the perceived direction of sound. This is possibly due to differences in auditory-visual integration between non-immersive and immersive display devices.

These results suggest that the increased immersion in VR may shift the cognitive workload and perceptual emphasis. It may also alter spatial mapping and error patterns. This fact shows the importance of platform-specific calibration in sound design to minimize localization inaccuracies.

4.3 Limitations of the Study

This study has several limitations that should be considered when interpreting the results. First, the participant pool itself. Although the sample size was statistically adequate, the participant pool lacked demographic diversity. The participants were mostly young and right-handed. The representation of older adults, left-handed individuals, and other diverse demographic groups was limited. This fact may introduce unintentional biases in sound localization accuracy. Future studies should consider a more diverse demographic sample, including left-handed users, older adults, and individuals with varied levels of VR familiarity, to improve generalizability and uncover subgroup-specific effects.

Another limitation relates to the cognitive demands placed on participants during immersive trials. Head-mounted displays such as the Gear VR often introduce a higher cognitive load due to the need for continuous multisensory integration, increased sensorimotor engagement, and novelty effects for inexperienced users. These factors can contribute to perceptual fatigue or attentional drift, especially in extended sessions. Although all participants completed the same number of trials, a formal cognitive load measurement tool (e.g., NASA-TLX or dual-task paradigms) was not implemented in this study to quantify mental effort. Future work could benefit from integrating such assessments to distinguish between perceptual limitations and cognitive strain.

The hardware characteristics of the devices used may have influenced performance. While enabling immersive interaction, the Gear VR has a relatively narrow field of view, limited motion tracking fidelity, and potential for physical discomfort such as motion sickness or neck strain. These limitations may impair sound localization accuracy or influence behavioral strategies (e.g., restricted head movement). In contrast, desktop displays may offer more consistent visual references but lack the naturalistic embodiment provided by VR, potentially leading to a less intuitive spatial frame of reference. The study did not control for perceptual latency, rendering fidelity, or user posture. Future studies should examine and control these factors as well.

Also, although the study was conducted in a VE designed to mimic naturalistic conditions, the sound localization task was still constrained by experimental simplifications. For instance, background noise was absent, and sound sources were discrete and isolated. These conditions do not reflect real-world auditory scenes. Moreover, participants were stationary during trials, whereas real-world localization often involves movement. Incorporating more ecologically valid scenarios, such as ambient soundscapes, moving sources, or dynamic interaction, would strengthen future designs.

Lastly, due to the COVID-19 pandemic, recruitment and testing processes were adjusted. This may have introduced inconsistencies in data collection and participant engagement. Therefore, the measures to ensure uniformity across conditions were limited by external constraints.

Conclusions

In this study, a virtual environment was created to measure the sound localization accuracy of users. During the measurement process, 110 participants used a desktop display and the Gear VR. The tests were randomized, and each participant had to localize sounds accurately ten times on each platform.

The results of this study show how display devices and human characteristics affect sound localization in virtual environments. The main results include gender differences in localization accuracy; however, the use of the Gear VR decreased this gap between the results. Previous VR experience improved performance in virtual spaces. The significant differences in inaccurate localization proportions between PC and VR show the influence of immersion level and interaction methods. These results contribute to optimizing sound localization features in virtual environments. By doing so, virtual spaces can be made more accessible and effective for users with various human characteristics. According to the results, future research should address the identified limitations to enhance generalizability and further refine VR-based auditory designs.

Beyond the theoretical implications, the results can also offer recommendations for real-world VR system design. For instance, VR training applications in several fields, such as emergency response, military simulation, and aviation, often rely on fast and accurate auditory cues. If specific display types degrade localization accuracy or increase cognitive load, system designers may need to adapt interface strategies such as enhancing spatial audio rendering or including supplementary visual indicators. Similarly, in healthcare and rehabilitation, selecting appropriate hardware can significantly impact therapeutic outcomes. The results support the idea that device-specific considerations are not only technical but also influence user performance and experience.

Acknowledgement

This work has been implemented by the TKP2021-NVA-10 project with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the 2021 Thematic Excellence Programme funding scheme.

The research presented in this paper was supported by the Hungarian Research Network (HUN-REN) and was carried out within the HUN-REN Cognitive Mapping of Decision Support Systems research group.

References

[1] Blauert, J.: Spatial hearing: Psychophysics of human sound localization. London, England: MIT Press, 1997

[2] Wogalter, M. S.: Handbook of Warnings, Lawrence Erlbaum and Associates. Mahwah, New Jersey, 2006

[3] Haas, E. C., van Erp, J. B. F.: Multimodal warnings to enhance risk communication and safety, Safety Science, Vol. 61, pp. 29-35, 2014

[4] Middlebrooks, J. C.: Sound localization. In The Human Auditory System - Fundamental Organization and Clinical Disorders, Elsevier, 2015, pp. 99-116

[5] Zahorik, P., Brungart, D. S., Bronkhorst, A. W.: Auditory distance perception in humans: A summary of past and present research. Acta Acustica, Vol. 91, pp. 409-420, 2005

[6] Zahorik, P.: Assessing auditory distance perception using virtual acoustics. The Journal of the Acoustical Society of America, Vol. 111, pp. 1832-1846, 2002

[7] Erdős, F., Németh, R., Firuza, B.: Virtual Teamwork in Gamified 3D Environment. Infocommunications Journal, Vol. 15, pp. 15-20, 2023

[8] Wersényi, G.: Musings on the Metaverse and on how the Internet of Digital Reality encompasses future developments of the Internet. Infocommunications Journal, Vol. 1, pp. 21-28, 2023

[9] Katona, J.: A review of human–computer interaction and virtual reality research fields in cognitive InfoCommunications. Applied Sciences, Vol. 11, p. 2646, 2021

[10] Drettakis, G., Roussou, M., Reche, A., Tsingos, N.: Design and evaluation of a real-world Virtual Environment for architecture and urban planning. Presence, Vol. 16, pp. 318-332, 2007

[11] Burdea, G., Coiffet, P.: Virtual Reality Technology. John Wiley & Sons, Nashville, TN, 2003

[12] Baranyi, P., Csapo, A., Sallai, G.C. Cognitive Infocommunications (CogInfoCom) Springer: Berlin/Heidelberg, Germany, 2015

[13] Kövecses-Gősi, V. Cooperative learning in VR environment. Acta Polytechnica Hungarica, Vol. 15, pp. 205-224, 2018

[14] Dæhlen, A., Heldal, I., Katona, J.: Towards Developing an Immersive Virtual Reality Applications for Supporting Vision Screening – A User Study. Journal of Applied Technical and Educational Sciences, Vol. 12, No. 4, 2023

[15] Faiqotuzzulfa, F., Putra, S. A.: Virtual reality's impacts on learning results in 5.0 education: A meta-analysis. International Transactions on Education Technology (ITEE), Vol. 1, pp. 10-18, 2023

[16] Ansari, S. Z. A., Shukla, V. K., Saxena, K., Filomeno, B.: Implementing virtual reality in entertainment industry. In Cyber Intelligence and Information Retrieval. Springer Singapore, Singapore, 2022, pp. 561-570

[17] Naceri, A., Chellali, R., Hoinville, T.: Depth perception within peripersonal space using head-mounted display. Presence, Vol. 20, pp. 254-272, 2011

[18] Guzsvinecz, T., Orbán-Mihálykó, É., Perge, E., Sik-Lányi, C.: Analyzing the spatial skills of university students with a virtual reality application using a desktop display and the gear VR. Acta Polytechnica Hungarica, Vol. 17, pp. 35-56, 2020

[19] Hug, D.: How do you sound design? An exploratory investigation of sound design process visualizations. In Proceedings of ACM Audio Mostly, 2020, pp. 114-121

[20] Storms, R. L., Zyda, M. J.: Interactions in perceived quality of auditory-visual displays. Presence, Vol. 9, pp. 557-580, 2000

[21] Turner, A., Berry, J., Holliman, N.: Can the perception of depth in stereoscopic images be influenced by 3D sound? In SPIE Proceedings, 2011

[22] Burke, J. L., Prewett, M. S., Gray, A. A., Yang, L., Stilson, F. R. B., Coovert, M. D., Elliot, L. R., Redden, E. Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. In Proceedings of the 8th international conference on Multimodal interfaces, 2006, pp. 108-117

[23] Hancock, P. A., Mercado, J. E., Merlo, J., Van Erp, J. B. F. Improving target detection in visual search through the augmenting multi-sensory cues. Ergonomics, Vol. 56, pp. 729-738, 2013

[24] Steadman, M. A., Kim, C., Lestang, J-H., Goodman, D. F. M., Picinali, L. Short-term effects of sound localization training in virtual reality. Scientific Repports, Vol. 9, p. 18284, 2019

[25] Yong, S., Wang, H. C.: Using spatialized audio to improve human spatial knowledge acquisition in virtual reality. In Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion, 2008, pp. 1-2

[26] Cullen, B., Galperin, D., Collins, K., Kapralos, B., Hogue, A: The effects of audio on depth perception in S3D games. In Proceedings of the 7th audio mostly conference: a conference on interaction with sound, 2012, pp. 32-39

[27] Guzsvinecz, T., Szabo, M., Halmosi, B., Sik-Lanyi, C. The Virtual Reality Sound-based Spatial Orientation Project. In Proceedings of the 12th IEEE International Conference on Cognitive InfoCommunications – CogInfoCom 2021, pp. 11-16

[28] Berger, M. Bill, R.: Combining VR visualization and sonification for immersive exploration of urban noise standards. Multimodal Technol. Interact., Vol. 3, No. 2, p. 34, 2019

[29] Wang W., Xu N., Dang S., Hong X., Qu, J.: Induction mechanism of auditory-assisted vision for target search localization in mixed reality (MR) environments, Aerospace, Vol. 9, No. 7, p. 340, 2022

[30] Moraes A. N., Flynn R., Hines A., Murray, N.: The role of physiological responses in a VR-based sound localization task, IEEE Access, Vol. 9, pp. 122082-122091, 2021

[31] Tai T.-Y., Chen, H. H.-J.: The impact of immersive virtual reality on EFL learners' listening comprehension, Journal of Educational Computing Research, Vol. 59, No. 7, pp. 1272-1293, 2021

[32] Reissmüller S.: Effects of sensory stimulation in Virtual Reality. TU Wien, 2024

[33] Baranyi, P., Csapo, A.: Definition and synergies of cognitive infocommunications, Acta Polytechnica Hungarica, Vol. 9, pp. 67-83, 2012

[34] Baranyi, P., Berki, B., Csapó, Á. B.: Concepts of Cognitive Infocommunications, Infocommunications Journal, vol. 16, Joint Special Issue on Cognitive Infocommunications and Cognitive Aspects of Virtual Reality, pp. 37-48, 2024

[35] Sallai, G. The cradle of cognitive infocommunications, Acta Polytechnica Hungarica, Vol. 9, No. 1, pp. 171-181, 2012

[36] Sudár, A., Horváth, I.: Design suggestions for digital workflow oriented desktop VR spaces, Infocommunications Journal., Vol. 16, Joint Special Issue on Cognitive Infocommunications and Cognitive Aspects of Virtual Reality, pp. 119-127, 2024

[37] Kővári, A.: Model of the Internet of Digital Education and its links to VR, Infocommunications Journal, Vol. 16, Joint Special Issue on Cognitive Infocommunications and Cognitive Aspects of Virtual Reality, pp. 76-83, 2024

[38] Guzsvinecz, T., Szűcs, J.: Textual Analysis of Virtual Reality Game Reviews, Infocommunications Journal, Vol. 16, Joint Special Issue on Cognitive Infocommunications and Cognitive Aspects of Virtual Reality, pp. 84-91, 2024