




## Article

# The Accuracy of Dynamic Sound Source Localization and Recognition Ability of Individual Head-Related Transfer Functions in Binaural Audio Systems with Head Tracking

Vedran Planinec <sup>1</sup>, Jonas Reijniers <sup>2</sup>, Marko Horvat <sup>1,\*</sup>, Herbert Peremans <sup>2</sup> and Kristian Jambrošić <sup>1</sup>

<sup>1</sup> Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia; vedran.planinec@fer.hr (V.P.); kristian.jambrosic@fer.hr (K.J.)

<sup>2</sup> Department of Engineering Management, Universiteit Antwerpen, 2000 Antwerp, Belgium; jonas.reijniers@uantwerpen.be (J.R.); herbert.peremans@uantwerpen.be (H.P.)

\* Correspondence: marko.horvat@fer.hr

**Abstract:** The use of audio systems that employ binaural synthesis with head tracking has become increasingly popular, particularly in virtual reality gaming systems. The binaural synthesis process uses the Head-Related Transfer Functions (HRTF) as an input required to assign the directions of arrival to sounds coming from virtual sound sources in the created virtual environments. Generic HRTFs are often used for this purpose to accommodate all potential listeners. The hypothesis of the research is that the use of individual HRTF in binaural synthesis instead of generic HRTF leads to improved accuracy and quality of virtual sound source localization, thus enhancing the user experience. A novel methodology is proposed that involves the use of dynamic virtual sound sources. In the experiments, the test participants were asked to determine the direction of a dynamic virtual sound source in both the horizontal and vertical planes using both generic and individual HRTFs. The gathered data are statistically analyzed, and the accuracy of localization is assessed with respect to the type of HRTF used. The individual HRTFs of the test participants are measured using a novel and efficient method that is accessible to a broad range of users.

**Keywords:** virtual sound source localization; acoustic signal processing; individual HRTF; binaural system with head tracking



**Citation:** Planinec, V.; Reijniers, J.; Horvat, M.; Peremans, H.; Jambrošić, K. The Accuracy of Dynamic Sound Source Localization and Recognition Ability of Individual Head-Related Transfer Functions in Binaural Audio Systems with Head Tracking. *Appl. Sci.* **2023**, *13*, 5254. <https://doi.org/10.3390/app13095254>

Academic Editor: Yat Sze Choy

Received: 22 March 2023

Revised: 12 April 2023

Accepted: 21 April 2023

Published: 23 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Binaural audio systems are used to create an immersive and realistic listening experience, providing listeners with a sense of spatial presence and directionality. With the rise of virtual and augmented reality technologies (VR and AR), binaural audio systems have become even more significant, as they play a crucial role in creating a convincing sensory experience [1]. Binaural audio uses Head-Related Transfer Functions (HRTF), which are essential for localizing sound in space and creating a sense of spatial awareness [2]. In virtual and augmented reality, binaural audio systems are used to create a realistic soundscape that matches the user's visual environment, providing a more immersive and engaging experience compared to using only visual cues [3]. From gaming and entertainment [4] or medicine [5] to education and training [6], binaural audio systems and spatial audio are revolutionizing the way we interact with virtual and augmented environments.

While generic HRTFs have been widely used in the field of spatial audio for creating a more accurate VR/AR experience [7], individual or personal HRTFs have the potential to personalize the experience even more. Furthermore, the use of generic HRTF can potentially lead to limitations in 3D audio perception in VR [8]. One of the potential principal benefits of utilizing individual HRTF as opposed to generic HRTF is the superior localization of sound sources within the frontal hemisphere, which may result in a higher degree of precision when employing an individual HRTF set that has been measured for a particular listener [9].

The acquisition of individual HRTF for a particular person needs to be performed using one of the many methods and measurement setups available [10]. Acquiring individual HRTF typically requires multiple exponential sweeps that employ a circular arc segment comprising many loudspeakers placed in an anechoic chamber. The arc can be maneuvered around a fixed central point where the subject is positioned. The subject wears a pair of in-ear microphones to measure their individual HRTF using such a setup [11–14]. Although the described conventional method of acquiring individual HRTF is still prevalent, recent years have seen the development of novel approaches that enable the acquisition of individual HRTF for a broader range of users without the need for complex measurement setups or an anechoic chamber. These new methods include techniques such as 3D scanning and modeling of individuals through structured-light/3D scanners or high-resolution pictures/video footage [15–17]. Another group of methods uses a single-speaker system placed in a typical room, and the subject wears a head-tracking unit equipped with Inertial Measurement Unit (IMU) sensors, implemented as a dedicated head-tracking unit or a smartphone [18–20]. For the purpose of this research, the methodology described in [18,21] was used to obtain individual HRTF for the subjects who took part in experimental measurements.

The primary objective of this study was to conduct experiments designed to test the hypothesis that the utilization of individual HRTF enhances the accuracy of localization of virtual sound sources, thus providing a more realistic and natural experience for the users compared to the situation when generic HRTF or the individual HRTF of another person are used. For this purpose, a novel methodology has been implemented for experiments that allow the hypothesis to be tested in highly dynamic virtual scenarios that emulate both real-life environments and virtual ones.

The presented novel methodology was also used to confirm that the localization accuracy in the horizontal plane will not be substantially affected by using non-individual HRTF instead of individual ones due to more robust additional interaural time and intensity difference cues, whereas the accuracy in the vertical plane will be significantly improved if an individual HRTF set is used in binaural synthesis.

## 2. Materials and Methods

### 2.1. Individual HRTF Measurement

The process of obtaining individual HRTF was conducted in accordance with [18], with the exception that the measurements were taken in an acoustically isolated small anechoic chamber. Before the measurement began, participants were given an oral explanation about HRTFs and the process of recording their own individual HRTF. The participants were then seated on a rotary chair in the anechoic chamber. The exact position was chosen to ensure that the listener's head is placed on-axis, i.e., in front of a single-driver loudspeaker at a distance of 1.5 m (as seen in Figure 1).

A head tracker was mounted onto the participant's forehead to measure the orientation data of the head. Shape-adjustable in-ear microphones were plugged into the subjects' ear canals to capture the exponential frequency sweeps emitted by the loudspeaker during the measurement, as in any acoustic HRTF measurement setup. Unlike in a standard HRTF measurement where the subject is obliged to remain motionless, the subjects now must move their heads with respect to the speaker to sample all the directions themselves. Indeed, they were instructed to sample the directions of arrival in the entire solid angle of  $4\pi$  in 15 min, with as few gaps as possible. To achieve this, they were also allowed to rotate their whole body while sitting on the chair so that the directions of arrival from the side and from behind could be sampled as well. To monitor their progress, they were presented with visual feedback on the directions that had already been sampled in the form of points on a sphere around their heads. The time remaining until the end of the measurement was displayed as well. This feedback was shown on two different computer screens placed on opposite sides of each other to maximize visibility during a full turn on the swivel chair. After the measurement, an additional system calibration measurement was carried out. The

measured data (IMU and audio recording) and the calibration data were then processed to produce the individual HRTF in Spatially Oriented Format for Acoustics (SOFA) [22].

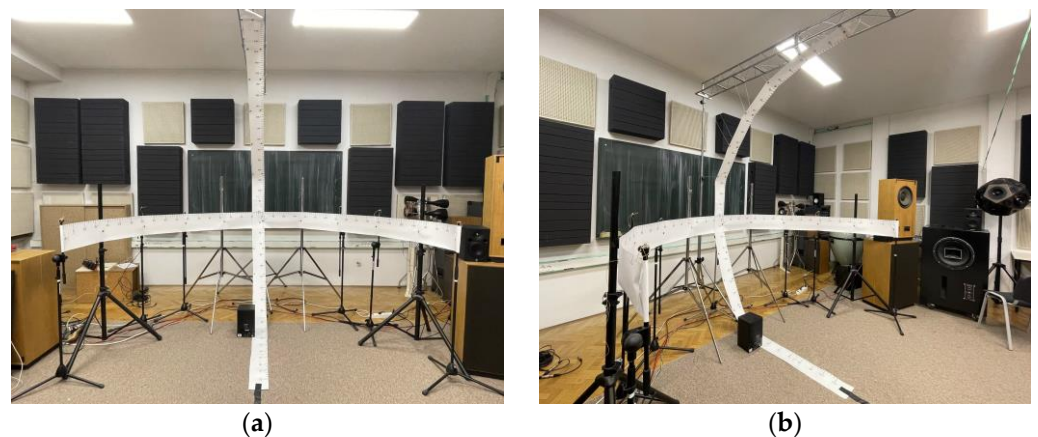


**Figure 1.** HRTF measurement in the anechoic chamber: (a) back view of a participant; (b) front view of a participant.

## 2.2. Measurement of Localization Accuracy

The binaural sound system designed for use in the localization accuracy experiment was implemented using a high-quality open-type headphone set, a high-quality external sound card, a PC, a wired head-tracker with a PC application for sending tracking data via OSC protocol [23], REAPER v6.75 as the digital audio workstation (DAW) software of choice [24], and a Sparta Binauralizer VST plug-in [25,26].

Before taking part in the experiment, the participants were given an informed consent form as well as written information and instructions about the research and the experiment. Both the headphone set and the head tracker were securely placed on the participant's head. To determine the direction of sound arrival, participants were seated on a rotating chair and allowed to freely pivot their bodies and heads. To localize the sound source and establish its azimuthal and elevation directions, two paper tapes were stretched symmetrically over a 180° arc around the listening position in the horizontal and vertical planes, with markings of azimuthal and elevation angles in 1° resolution (as seen in Figure 2). The radius of the paper arc was 1.5 m, and the precision of the markings was checked with the digital protractor.



**Figure 2.** Horizontal and vertical paper tapes used for determining the azimuth and the elevation of the sound source: (a) front view; (b) side view.

To avoid negative values of azimuth and elevation, which would be more difficult for the subjects to read, the frontal direction was assigned an azimuth and elevation of 90°, thus yielding a range from 0° to 180° for both parameters. The participants were asked to

determine the starting and final azimuth (or elevation) of a virtual sound source moving in the horizontal (or vertical) plane within the chosen range of azimuth (elevation) angles. The sound stimulus employed in this experiment was a single knock on a piece of wood that repeated itself at a rate of 3.53 knocks per second. The total duration of the stimulus was randomized, ranging from 8 to 12 s. The direction of arrival of the sound source continuously changed from a starting value to an ending value at a non-consistent angular velocity (ranging from  $4^\circ$  to  $10^\circ$  per second). The angular velocity of the virtual sound source was randomized (by no more than  $3^\circ$  per second from the starting angular velocity) to avoid automatic turning of the head after adjusting to the constant speed of the sound source (in the case of a constant angular velocity). As an additional measure implemented to avoid automatic turning of the head, in some cases the sound stimulus in the horizontal plane was allowed to have a turning azimuth as the position at which the virtual source changes its direction of movement, i.e., starts moving back towards the starting position. The vertical and horizontal planes were tested independently, and the participants were informed which paper tape they should focus on before every individual measurement. The participants were instructed to verbally indicate the initial direction of the virtual sound source immediately upon localizing it, as well as the final direction from which they perceived the sound following the conclusion of the sound stimulus (as seen in Figure 3).



**Figure 3.** A participant with head-mounted binaural system elements on a rotating chair, tracking the virtual sound stimulus.

Before the actual testing began, two test examples were presented to each participant to verify the functionality of the binaural system and ensure that the participants understood the experimental protocol.

In total, each participant listened to 18 test cases and indirectly evaluated the effectiveness of three different HRTFs. The three HRTFs under investigation included the participant's own individual HRTF, a generic HRTF obtained from the Neumann KU100 dummy head [27], and the individual HRTF of another person, obtained from a single participant who had his individual HRTF measured but did not participate in the experiment. The SOFA file containing the generic HRTF was the only one with 2702 directions defined, while both individual HRTFs had 1460 directions defined ( $3^\circ$  resolution). While a higher number of directions measured for the generic HRTF might imply a potential advantage of the generic HRTF over the individual ones regarding the localization accuracy in the horizontal plane, it is unlikely to have an impact on the localization accuracy in the vertical plane.

To minimize the possibility of developing a bias, measurements taken in the horizontal and vertical planes were systematically alternated every three measurements. Furthermore, to mitigate potential confounding effects, the three HRTFs were randomly rearranged after every three-measurement segment.



### 2.3. The Ability of the Listeners to Recognize Their Own Individual HRTF

This experiment involved the assessment of various HRTFs by the participants, who were given the task of selecting their own individual HRTF from a group of four different HRTFs by listening to sound samples that were binaurally encoded using these HRTFs. The objectives of this experiment were two-fold: firstly, to serve as a training exercise for the participants, aimed at familiarizing them with head-tracked virtual audio and sensitizing them to pertinent acoustic cues. Secondly, the study aimed to indicate the potential benefits of individual HRTFs in a virtual audio system for each subject. This advantage may translate to improved performance in the primary focus of this research, namely, the accuracy of sound localization. Therefore, this experiment was performed before the localization accuracy experiment.

The experimental setup involved seating the participants in front of a computer and equipping them with headphones and a head tracker. The experiment was conducted using a C++ program that utilizes the 3DTI toolkit [28].

Before the experiment began, the precise orientation of the head tracker relative to both the head and the world frame needed to be determined by asking the participants to nod their heads for “yes” and “no”. The experimental procedure consisted of 10 trials in total, wherein a virtual sound source was positioned directly in front of the participant, with its direction stabilized and updated in real-time based on head-tracking data. The participants were given the option to select between four randomized HRTFs by pressing a corresponding number on the keyboard and were allowed to change their choice for as long as they saw fit. The participants were instructed to choose the preferred HRTF according to the following criteria: “Choose the HRTF which you think is yours, considering the natural feeling of the sound, the externalization of the sound, and it should sound like you are listening with your own ears, and not someone else’s”. The “natural feeling” was additionally defined to the participants as the perception of a sound source that appears consistently in the same, well-localized direction, regardless of head rotation. The same audio files were played repeatedly throughout the trial, with the switching of the HRTF occurring instantaneously upon the participant’s command via the keyboard. The trial concluded once the participant communicated their preferred HRTF via the keyboard. Two distinct stimuli were presented, alternating between consecutive trials:

1. A woman reading a technical text in English;
2. A song by Ozark Henry.

The same set of four randomized HRTFs was employed in each trial. The set consisted of the participant’s own individual HRTF and the individual HRTFs of three other people who did not partake in the experiment. These three HRTFs were identical for all participants.

### 2.4. System Latency Measurement

Latency is regarded as one of the critical concerns in virtual audio systems [29]. To ensure that accurate results are obtained from the designed experiments, it was necessary for the system latency of the binaural audio system to be sufficiently low. One of the key factors was the use of a high-quality wired head tracker, as the performance of the head tracker itself has a considerable influence on accuracy, precision, and low latency in binaural audio systems [30]. Any experiment that utilizes binaural audio systems requires the latency of the system to be as low as possible, optimally below 50 ms. Although a theoretical latency value could be calculated, a more precise experimental approach was employed so that the actual value could be determined.

The binaural audio system was mounted on a dummy head, along with a simple electrical circuit for light detection consisting of a 9V battery, a light-emitting diode, and a high-quality photoresistor. A specialized SOFA file was implemented in the binaural audio system, where a Dirac impulse with amplitude 0.5 was set up for the frontal HRTF directions (within 6° of great angle error), while amplitude 0 was set up for all other directions. A looping 10-kHz sinusoidal tone was played into the binaural system, with the dummy head positioned in the frontal direction (i.e., where the 10-kHz tone was audible).

A laser pointer was directed toward the photoresistor, which decreased the element's resistance, causing the voltage on the photoresistor to rise. With the sudden rotation of the dummy head, the laser beam moved away from the photoresistor, leading to a drop in voltage. At the same time, the head-tracker of the binaural system detected movement of the dummy head, causing the system to change azimuth and change HRTF to amplitude 0 (i.e., the sound was no longer audible). The latency was subsequently calculated as the time difference between the exact moment the voltage began to rise on the photoresistor and the moment the audio output of the binaural system switched from audible to inaudible.

A two-input sound card was utilized to record both the audio output of the binaural system and the voltage change on the photoresistor. Modern external sound cards typically have a capacitor in series with the input to block DC voltages from entering the sound card. Therefore, the required voltage drop could be observed as a starting point for the change in the AC recorded signal. To maximize the voltage drop effect, the measurement was conducted in a controlled environment devoid of external light sources.

The latency of the individual HRTF selection setup was measured to be 35 ms, which was deemed sufficiently low for the progression of the experiment described in Section 2.3.

Ten latency measurements were conducted on the binaural system setup designed to test localization accuracy. The obtained latency values spanned from 69 to 78 ms, with an average result of 73 ms. Given that the time interval between 2 successive knocks of the sound stimulus was 283 ms and that the latency was nearly 4 times shorter than the mentioned interval, the obtained latency was deemed sufficiently low for the progression of the localization accuracy experiment.

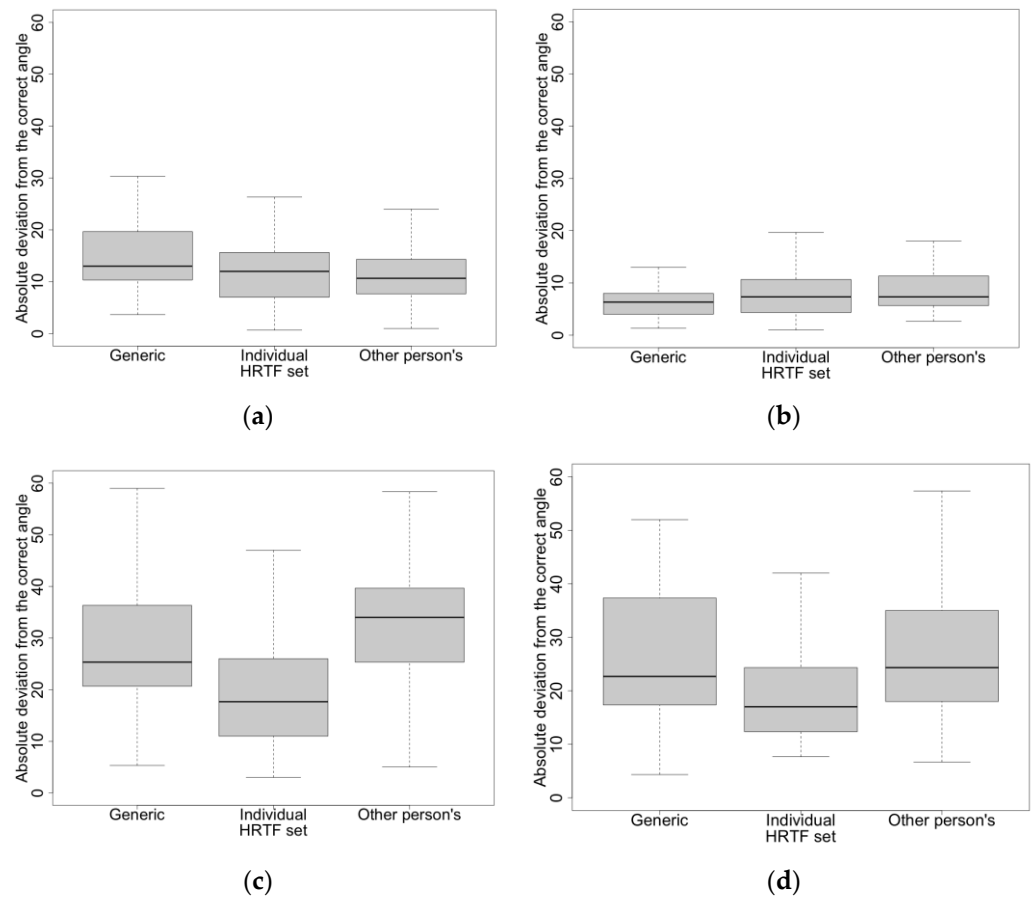
### 3. Results

The following subsections report the results of statistical analyses performed on raw data collected during the experiments and on adjusted raw data. In the statistical analysis, all decisions on statistical significance were made at the 0.05 significance level. The statistical analyses were performed in R [31].

#### 3.1. Localization Accuracy Results

The statistical analysis is based on repeated measures of analysis of variance (ANOVA) [32] and was performed according to [33]. Since there were 25 independent participants with multiple measured data points, their results with the same parameters (HRTF, vertical or horizontal plane, starting or ending angle) were averaged as independent data points. In total, each participant contributed 12 data points for the analysis. For each of the three HRTFs, there were horizontal or vertical starting and ending perceived sound source directions represented by their azimuth or elevation angles, respectively. The obtained data that was statistically analyzed were the absolute (non-negative) values of the deviation from the correct azimuth (or elevation) starting and ending directions of the virtual sound source. If the occasional missing data arose from the participant's unsureness about the starting or ending angle of the virtual sound source, the average absolute deviation for the corresponding data point was computed without the missing value by utilizing other available deviation data from the same data point. There was always at least one selected angle for each data point for each participant, resulting in no missing data from the independent data points for all the participants, which was an important factor regarding the proper usage of repeated measures in ANOVA.

Since there were a small number of outliers (less than two outliers for every 25-point data set), the mean/median imputation technique was employed to adapt outlier values for ANOVA. Normality testing using Shapiro–Wilk's test was performed on all data sets, demonstrating that the data are normally distributed. Finally, Mauchly's test of sphericity was used to verify that the null hypothesis of equivalent variances of differences was fulfilled. Box and whisker plots were generated for the absolute deviation from true horizontal and vertical starting and ending points, as presented in Figure 4.



**Figure 4.** Box and whisker plots for the absolute deviation from a true angle as the measure of localization accuracy, depending on the used HRTF, for: (a) the horizontal starting angle; (b) the horizontal ending angle; (c) the vertical starting angle; and (d) the vertical ending angle of the virtual sound source.

ANOVA results were calculated as follows:

- For the horizontal starting angle:  $F(48,2) = 7.052, p = 0.002$ ;
- For the horizontal ending angle:  $F(48,2) = 3.223, p = 0.049$ ;
- For the vertical starting angle:  $F(48,2) = 10.113, p = 0.000216$ ;
- For the vertical ending angle:  $F(48,2) = 4.542, p = 0.016$ .

The data obtained from the experiment indicates the existence of statistically significant differences for all four investigated parameters regarding the HRTF that was used. To further assess the performance of individual HRTFs compared to other HRTFs, a pairwise comparison was performed, and the results are presented in Table 1.

**Table 1.** Pairwise comparison of differences between HRTF sets.

	Gen&Oth Pair <sup>1</sup>		Ind&Gen Pair <sup>2</sup>		Ind&Oth Pair <sup>3</sup>	
	<i>p</i> -Value	adj. <i>p</i> -Value <sup>4</sup>	<i>p</i> -Value	adj. <i>p</i> -Value <sup>4</sup>	<i>p</i> -Value	adj. <i>p</i> -Value <sup>4</sup>
Horizontal starting angle	0.002	0.006	0.005	0.015	0.894	1.000
Horizontal ending angle	0.031	0.094	0.086	0.256	0.623	1.000
Vertical starting angle	0.214	0.642	0.002	0.007	0.0006	0.002
Vertical ending angle	0.674	1.000	0.007	0.022	0.016	0.050

<sup>1</sup> Comparison between generic HRTF and the individual HRTF of another person. <sup>2</sup> Comparison between the participant's own individual HRTF and generic HRTF. <sup>3</sup> Comparison between the participant's own individual HRTF and the individual HRTF of another person. <sup>4</sup> *p*-value adjusted with the Bonferroni correction.

The outcomes concerning the horizontal starting angle indicate that there is no statistically significant difference in absolute deviation from the true angle between one’s own individual HRTF and the individual HRTF of another person. However, a statistically significant difference is observed between both of these cases and the case when generic HRTF is used.

For the horizontal ending angle, there was no statistically significant difference between HRTFs.

The results obtained from the pairwise comparison indicate a significantly lower absolute deviation from true vertical starting and ending angles when one’s own individual HRTF is used compared to the remaining two HRTFs.

Additionally, the measurement results acquired from the participants were analyzed for the correct determination of the direction in which the virtual sound source moved. If the angular value of the sound source was increasing from the starting to the ending angle, it was assigned the value “1”, and if it was decreasing, it was assigned the value “0”. This was performed for both the programmed values to determine the correct movement direction and the measurement results. The results are presented in Table 2.

**Table 2.** Correct determination of the direction in which the sound source moves (in percent).

	Generic HRTF	One’s Own Individual HRTF	Individual HRTF of Another Person
Horizontal plane	98.67%	98.67%	100%
Vertical plane w/m.d. <sup>1</sup>	40%	81.33%	50.67%
Vertical plane w/o m.d. <sup>2</sup>	44.12%	83.56%	55.07%

<sup>1</sup> Results with missing data included in incorrect observations. <sup>2</sup> Results with missing data excluded from incorrect observations.

Although the missing data did not affect the absolute deviation calculations because of the independent data point averaging, it should be considered for the correct determination of the direction calculations. When the missing data arose, there were two possible ways of handling it: defining the direction determination as incorrect or excluding the missing data from further analysis. In the case of a participant’s unsureness about where the sound source started or ended (missing data), the direction could not be defined with absolute certainty; therefore, the exclusion of the missing data is a more pertinent approach. There was no missing data for the horizontal plane.

A chi-square ( $X^2$ ) test was conducted to test whether the scores obtained for correct and incorrect determination of the direction of movement in the vertical plane carry any statistically significant difference from the values expected from purely guessing the direction by random chance. This was particularly important regarding the generic HRTF since the correct determination percentage was slightly below 50%. The results of the chi-square test ( $p = 0.08$  for missing data included in incorrect determinations,  $p = 0.33$  for missing data excluded from incorrect determinations) imply that no statistically significant difference exists in this case, i.e., the determination of direction in the vertical plane using generic HRTFs is essentially based on guessing. A binomial test was also conducted, and it confirms the results of the chi-square test ( $p = 0.3961$ ). The presented results demonstrate that using one’s own individual HRTF leads to considerably higher accuracy when determining the direction in which the sound source moves in the vertical plane.

### 3.2. The Success in Selecting One’s Own Individual HRTF

The number of trials in which participants selected their own individual HRTF was tabulated for each of the 25 participants. Subsequently, a binomial test was conducted to examine whether this count was significantly greater than the expected value of 2.5 for 10 trials (equivalent to random chance). The success rate of choosing one’s own individual HRTF was significantly higher ( $p < 0.05$ ) for five subjects. Upon pooling the data of all participants, the overall success rate was calculated to be 87 out of 250 trials. As a result, the



likelihood of participants selecting their own HRTF over the other ones was significantly higher than random chance ( $p$ -value < 0.001).

#### 4. Discussion

The results of the statistical analysis confirm the hypothesis that the use of one's own individual HRTF results in improved accuracy in localizing virtual sound sources in the vertical plane. This is particularly evident in determining the starting angle of the sound stimulus in the vertical plane. As the initial orientation of the participant's head requires them to look directly at the intersection of horizontal and vertical paper strips, the employment of individual HRTF furnished the participant with more precise localization information. Additionally, the data that indicates the determination of sound source movement in the vertical plane (83.56% correctly determined directions of movement for one's own individual HRTF, compared to 55.07% for the HRTF of another person and 44.12% for generic HRTF) shows that using one's own individual HRTF enhances the accuracy of dynamic localization. Consequently, this result indicates that using one's own HRTF can lead to a more realistic perception of the virtual auditory environment and foster a more immersive spatial awareness.

The results of the statistical analysis regarding the localization accuracy in the horizontal plane indicate that no statistically significant improvement is obtained by using one's own individual HRTF compared to the HRTF of another person. The use of generic HRTF when determining the horizontal starting angle leads to the poorest localization accuracy, which is significantly lower compared to the other tested HRTFs. The box and whisker plot that indicates the accuracy of determining the horizontal ending angle shows that the use of the generic HRTF has yielded slightly better results than other HRTFs, but the improvement was not statistically significant. A possible reason for this is the higher resolution of the generic HRTF, with 2702 defined directions compared to the 1460 defined directions for the individual HRTFs. To examine this issue (if relevant), future work will consider the same density of direction points for all HRTFs.

A possible reason for the lack of improvement in horizontal plane localization by using one's own individual HRTF could be that the primary cues for horizontal sound localization, such as interaural time differences (ITD) and interaural level differences (ILD), are primarily determined by the ear canal and not significantly affected by individual head and pinna morphology [34]. Therefore, for sound source localization in the horizontal plane, generic HRTFs have the potential to be used without a significant loss (if any) in localization accuracy.

The results obtained from the experiment that tested the ability to recognize one's own individual HRTF show that the listeners as a group displayed a certain degree of ability to identify their own individual HRTF when their HRTF was grouped with three HRTFs of other people, which goes beyond random chance. In this experiment, the participants were asked only to select the HRTF that they believed to be their own. A more effective approach, to be considered in future work, would be to implement a ranking system whereby all four HRTFs that were used in the experiment are ranked in order of preference. This approach would be particularly useful in situations where participants may struggle to differentiate between their own HRTF and the specific HRTF of another person. Furthermore, an acknowledged limitation of the presented approach is that a reference sound signal could have been used. Although the use of the reference signal could imply more significant results in the presented experiment, there is a reason for not using it in the current setup. Since it would be necessary for the participants to remove their headsets for the reference signal, that could disrupt the calibration of the head tracker attached to the headphones. In future work, adequate use of the reference sound signal could be considered so that it would not influence the head tracker calibration.

Additionally, an analysis was performed to investigate the existence of a correlation between the results of the individual HRTF recognition experiment and the results of the localization experiment. However, the findings revealed no statistically significant

correlation. With the improvements implemented into the methodology of the HRTF recognition experiment in future work, the possibility of a correlation between the results of these two experiments could be revisited.

Concerning the latency of the binaural audio system used in the localization experiment, it is expected that the shortening of the latency time to below 50 ms may further improve the localization accuracy when one's own individual HRTFs are used. In future research, upgrades made to the binaural system setup could lead to lower latency, thus providing a solid base for new experiments, the results of which would help verify this premise.

Despite using a different method of measuring individual HRTF and a different localization methodology, the presented results agree with previous studies [35–37] that also showed statistically significant performance improvement when one's own individual HRTF is used. This finding goes in favor of confirming the claim that the use of one's own individual HRTF provides a more natural and immersive listening experience. One study, in particular, confirms improvements in sound source localization by using individual HRTF over generic, but the emphasis is put on the compelling performance of a generic HRTF set [38]. On the other hand, one study found no difference between using individual HRTF and generic HRTF in dynamic virtual environments [39], although their methodology and experimental design are considerably different from the ones presented in this research. The implications of these contradictory findings suggest that additional investigation might help to identify the most informative dynamic cues in order to enhance our comprehension of the variations between the experimental conditions, which may clarify the underlying factors contributing to the observed discrepancies.

## 5. Conclusions

In summary, this study presents a novel methodology for assessing the localization accuracy in binaural audio systems, with an emphasis on identifying possible improvements by using one's own individual HRTF instead of a generic HRTF or an individual HRTF of another person. Although the results show statistically significant improvements in accuracy for one's own individual HRTF in the vertical plane, the results also suggest the absence of universal statistically significant improvements in localization accuracy in the horizontal plane when using one's own HRTF compared to other HRTFs. The individual HRTF recognition experiment also found an indication of statistical significance for the ability to recognize one's own HRTF, but the methodology warrants further refinement in future work. Future studies will focus on comparing the effectiveness of the individual HRTF measurement method used in this research with other approaches for measuring individual HRTF.

**Author Contributions:** Conceptualization, V.P., K.J., J.R. and H.P.; methodology, V.P., K.J., J.R. and H.P.; software, V.P., J.R. and H.P.; validation, M.H. and K.J.; formal analysis, V.P. and J.R.; investigation, V.P. and J.R.; resources, K.J., M.H. and H.P.; data curation, V.P. and J.R.; writing—original draft preparation, V.P.; writing—review and editing, J.R., H.P., M.H. and K.J.; visualization, V.P.; supervision, K.J. and H.P.; project administration, K.J. and H.P.; funding acquisition, K.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Croatian Science Foundation (HRZZ IP-2018-01-6308, "Audio Technologies in Virtual Reality Systems for Auralization Applications (AUTAURA)") and by the Flemish Agency for Innovation and Entrepreneurship (VLAIO).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and the Ethical Code of the University of Zagreb and approved by the Ethics Committee of the Faculty of Electrical Engineering and Computing (protocol code EP-17-18, approved on 17 July 2018).

**Informed Consent Statement:** Informed consent was obtained from all participants involved in the study.

**Data Availability Statement:** The raw data collected for the purpose of this study are available upon request from the corresponding authors. Even though the data are anonymous, it were not made publicly available to maintain confidentiality and protect the privacy of the participants as a group.

**Acknowledgments:** The authors would like to thank Stjepan Šebek for his generous help with the statistics calculation and implementation. The authors also appreciate the help from all of the participants that were part of the experiments.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Middlebrooks, J.C. Sound localization. In *Handbook of Clinical Neurology*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 129, pp. 99–116. [CrossRef]
2. Wightman, F.L.; Kistler, D.J. Headphone simulation of free-field listening. I: Stimulus synthesis. *J. Acoust. Soc. Am.* **1989**, *85*, 858–867. [CrossRef] [PubMed]
3. Shilling, R.D.; Shinn-Cunningham, B. Virtual auditory displays. In *Handbook of Virtual Environment Technology*; Stanney, K., Ed.; Lawrence Erlbaum: Mahwah, NJ, USA, 2002; pp. 65–92.
4. Broderick, J.; Duggan, J.; Redfern, S. The Importance of Spatial Audio in Modern Games and Virtual Environments. In Proceedings of the 2018 IEEE Games, Entertainment, Media Conference (GEM), Galway, Ireland, 16–18 August 2018; pp. 1–9. [CrossRef]
5. Johnston, D.; Egermann, H.; Kearney, G. The Use of Binaural Based Spatial Audio in the Reduction of Auditory Hypersensitivity in Autistic Young People. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12474. [CrossRef] [PubMed]
6. Dede, C.; Jacobson, J.; Richards, J. Introduction: Virtual, Augmented, and Mixed Realities in Education. In *Virtual, Augmented, and Mixed Realities in Education*; Liu, D., Dede, C., Huang, R., Richards, J., Eds.; Springer: Singapore, 2017; pp. 1–16.
7. Berger, C.C.; Gonzalez-Franco, M.; Tajadura-Jiménez, A.; Florencio, D.; Zhang, Z. Generic HRTF May be Good Enough in Virtual Reality. Improving Source Localization through Cross-Modal Plasticity. *Front. Neurosci.* **2018**, *12*, 21. [CrossRef] [PubMed]
8. Jenny, C.; Reuter, C. Usability of Individualized Head-Related Transfer Functions in Virtual Reality: Empirical Study with Perceptual Attributes in Sagittal Plane Sound Localization. *JMIR Serious Games* **2020**, *8*, e17576. [CrossRef] [PubMed]
9. Møller, H. Fundamentals of Binaural Technology. *Appl. Acoust.* **1992**, *36*, 171–218. [CrossRef]
10. Li, S.; Peissig, J. Measurement of Head-Related Transfer Functions: A Review. *Appl. Sci.* **2020**, *10*, 5014. [CrossRef]
11. Møller, H.; Sørensen, M.F.; Hammershøi, D.; Jensen, C.B. Head-related transfer functions of human subjects. *J. Audio Eng. Soc.* **1995**, *43*, 300–321.
12. Richter, J.; Behler, G.; Fels, J. Evaluation of a Fast HRTF Measurement System. In Proceedings of the 140th International AES Convention, Paris, France, 4–7 June 2016; p. 9498.
13. Pausch, F.; Doma, S.; Fels, J. Hybrid multi-harmonic model for the prediction of interaural time differences in individual behind-the-ear hearing-aid-related transfer functions. *Acta Acust.* **2022**, *6*, 34. [CrossRef]
14. Carpentier, T.; Bahu, H.; Noisternig, M.; Warusfel, O. Measurement of a Head-Related Transfer Function Database with High Spatial Resolution. In Proceedings of the 7th Forum Acusticum, European Acoustics Association (EAA), Krakow, Poland, 7–12 September 2014.
15. Braren, H.S.; Fels, J. *A high-Resolution Individual 3D Adult Head and Torso Model for HRTF Simulation and Validation: HRTF Measurement*; Technical Report, Published under Creative Commons Attribution 4.0 License; RWTH Aachen: Aachen, Germany, 2020. [CrossRef]
16. Mäkitvirta, A.; Malinen, M.; Johansson, J.; Saari, V.; Karjalainen, A.; Vosough, P. Accuracy of Photogrammetric Extraction of the Head and Torso Shape for Personal Acoustic HRTF Modeling. In Proceedings of the Audio Engineering Society Convention 148, May 2020, Virtual Vienna, 2–5 June 2020; p. 10323.
17. Hoene, C.; Patino Mejia, I.S.; Cacerovschi, A. MySofa—Design Your Personal HRTF. In Proceedings of the 142nd International AES Convention, Berlin, Germany, 20–23 May 2017; p. 9764.
18. Reijniers, J.; Partoens, B.; Peremans, H. DIY Measurement of Your Personal HRTF at Home: Low-Cost, Fast and Validated. Convention e-Brief 399. In Proceedings of the 143rd International AES Convention, New York, NY, USA, 18–21 October 2017.
19. Gan, W.S.; Peksi, S.; He, J.; Ranjan, R.; Hai, N.D.; Chaudhary, N.K. Personalized HRTF Measurement and 3D Audio Rendering for AR/VR Headsets. In Proceedings of the Audio Engineering Society 142nd International Convention Committee Announced, Berlin, Germany, 20–23 May 2017.
20. Zandi, N.H.; El-Mohandes, A.M.; Zheng, R. Individualizing Head-Related Transfer Functions for Binaural Acoustic Applications. *arXiv* **2022**, arXiv:2203.11138. [CrossRef]
21. Reijniers, J.; Partoens, B.; Steckel, J.; Peremans, H. HRTF Measurement by Means of Unsupervised Head Movements with Respect to a Single Fixed Speaker. *IEEE Access* **2020**, *8*, 92287–92300. [CrossRef]
22. The Acoustics Research Institute of the Austrian Academy of Sciences. SOFA (Spatially Oriented Format for Acoustics). Available online: [https://www.sofaconventions.org/mediawiki/index.php/SOFA\\_\(Spatially\\_Oriented\\_Format\\_for\\_Acoustics\)](https://www.sofaconventions.org/mediawiki/index.php/SOFA_(Spatially_Oriented_Format_for_Acoustics)) (accessed on 17 March 2023).
23. Wright, M.; Freed, A. Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. In Proceedings of the International Computer Music Conference (ICMC), Thessaloniki, Greece, 25–30 September 1997; pp. 101–104.
24. COCKOS Inc. Reaper: Digital Audio Workstation. Rosendale, NY, USA, 2023. Available online: <https://www.reaper.fm> (accessed on 17 March 2023).

25. McCormack, L.; Politis, A. SPARTA & COMPASS: Real-Time Implementations of Linear and Parametric Spatial Audio Reproduction and Processing Methods. In Proceedings of the AES International Conference on Immersive and Interactive Audio, York, UK, 27–29 March 2019; p. 111.
26. McCormack, L.; Politis, A. Spatial Audio Real-Time Applications. Available online: [http://research.spa.aalto.fi/projects/sparta\\_vsts/](http://research.spa.aalto.fi/projects/sparta_vsts/) (accessed on 17 March 2023).
27. Neumann KU100 Operating Instructions. Technical Report. Available online: <https://www.manualslib.com/manual/110720/Neumann-Berlin-Dummy-Head-Ku-100.html> (accessed on 17 March 2023).
28. Cuevas-Rodríguez, M.; Picinali, L.; González-Toledo, D.; Garre, C.; de la Rubia-Cuestas, E.; Molina-Tanco, L.; Reyes-Lecuona, A. 3D Tune-In Toolkit: An Open-Source Library for Real-Time Binaural Spatialisation. *PLoS ONE* **2019**, *14*, e0211899. [[CrossRef](#)] [[PubMed](#)]
29. Kapralos, B.; Jenkin, M.; Miliotis, E. *Virtual Audio Systems in Presence Teleoperators & Virtual Environments*; MIT Press: Cambridge, MA, USA, 2008; Volume 17, pp. 527–549. [[CrossRef](#)]
30. Franček, P.; Jambrošič, K.; Horvat, M.; Planinec, V. The Performance of Inertial Measurement Unit Sensors on Various Hardware Platforms for Binaural Head-Tracking Applications. *Sensors* **2023**, *23*, 872. [[CrossRef](#)] [[PubMed](#)]
31. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021; Available online: <https://www.R-project.org/> (accessed on 17 March 2023).
32. Park, E.; Cho, M.; Ki, C.S. Correct use of repeated measures analysis of variance. *Korean J. Lab. Med.* **2009**, *29*, 1–9. [[CrossRef](#)] [[PubMed](#)]
33. Finnstats. Repeated Measures of ANOVA in R Complete Tutorial. Available online: <https://finnstats.com/index.php/2021/04/06/repeated-measures-of-anova-in-r/> (accessed on 17 March 2023).
34. Middlebrooks, J.C.; Green, D.M. Sound localization by human listeners. *Annu. Rev. Psychol.* **1991**, *42*, 135–159. [[CrossRef](#)] [[PubMed](#)]
35. Oberem, J.; Richter, J.-G.; Setzer, D.; Seibold, J.; Koch, I.; Fels, J. Experiments on localization accuracy with non-individual and individual HRTF comparing static and dynamic reproduction methods. *BioRxiv* **2020**, arXiv:31.011650. [[CrossRef](#)]
36. Wang, L.; Zeng, X.; Ma, X. Advancement of Individualized Head-Related Transfer Functions (HRTFs) in Perceiving the Spatialization Cues: Case Study for an Integrated HRTF Individualization Method. *Appl. Sci.* **2019**, *9*, 1867. [[CrossRef](#)]
37. Andersen, J.S.; Miccini, R.; Serafin, S.; Spagnol, S. Evaluation of Individualized HRTFs in a 3D Shooter Game. In Proceedings of the 2021 Immersive and 3D Audio: From Architecture to Automotive (I3DA), Bologna, Italy, 16–17 September 2021; pp. 1–10. [[CrossRef](#)]
38. Armstrong, C.; Thresh, L.; Murphy, D.; Kearney, G. A Perceptual Evaluation of Individual and Non-Individual HRTFs: A Case Study of the SADIE II Database. *Appl. Sci.* **2018**, *8*, 2029. [[CrossRef](#)]
39. Rummukainen, O.S.; Robotham, T.; Habets, E.A.P. Head-Related Transfer Functions for Dynamic Listeners in Virtual Reality. *Appl. Sci.* **2021**, *11*, 6646. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.