Audio Engineering Society

Conference Paper

Presented at the AES 2026 International Conference on
Audio for Virtual and Augmented Reality and Immersive Games
2026 June 30 – July 3, Paris, France

This paper was peer-reviewed as a complete manuscript for presentation at this conference. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Which Tracking Characteristics from an Audio Only VR Onboarding Predict the Best Performance HRTFs?

Max Væhrens[1], Stefania Serafin[2], Flemming Christensen[1], and Dorte Hammershøi[1]

[1]*Acoustics and Hearing, AI and Sound, Department of Electronic Systems, Aalborg University*
[2]*Department of Engineering Technology and Didactics, Technical University of Denmark*

Correspondence should be addressed to Max Væhrens (maxv@es.aau.dk)

ABSTRACT

This study is motivated by an ambition to determine the 'best'-matching HRTFs during an onboarding task for an audio-only virtual reality (VR) experience using a 'shooting down sound sources' task. The study is motivated by the needs of blind and visually impaired gamers, who may rely more crucially on accurate rendering of auditory spatial cues for succeeding in the audio-only VR experience. We present an exploratory study applying an experimental VR test platform that renders 'target' sound sources in a virtual environment and logs tracking characteristics of head, hand-held controller and body while participants localise and 'shoot' audible 'targets' that are visible (for task familiarisation) and invisible. Four game-relevant sound stimuli and three different HRTFs were tested across eight sessions on two separate days. In this study, we show data collected from fifteen seeing participants, which demonstrate an ability to localise the sound sources accurately. The tracking data suggests various search patterns (e.g. hemisphere swaps and direction reversals) associated with 'weak' localisation cues and possible ambiguities. The search patterns are likely all quantifiable via angular error, response time, path length, search directions, number of reversals, and search speed as determined from the tracking characteristics.

1 Introduction

Virtual reality (VR) games predominantly utilise visuals to render both game scenes and controller interaction. This dependency on visual rendering largely constitutes an accessibility barrier to blind and visually impaired (BVI) gamers. The long-term purpose of the present efforts is to further develop opportunities for audio-only VR.

Prior research indicates that in terms of the auditory rendering of game scenes, BVI individuals may be a more critical user population [1, 2, 3, 4, 5, 6, 7, 8] (however this hypothesis is also contested [9, 10]). Additionally, since the BVI typically use their hearing more often when navigating and identifying sources in their vicinity, they may be better able to discriminate between subtle differences in sound cues and room acoustics. Hence, we hypothesise that BVI gamers may rely more crucially on having individual head-related transfer functions (HRTFs) [11] for the auditory rendering in audio-only games.

The literature has established that individual HRTFs are superior for localising sound sources [11, 12, 13] and for speech understanding [14]. There are also reports

which suggest that specific sets of 'generic' HRTFs yield better localisation performance compared with other randomly selected HRTFs [15]. It has also been reported that training on non-individual HRTFs can improve performance [16, 17], although no training improvements have been observed [18]. Hence, it is difficult to know whether a specific HRTF implementation provides only minor improvements or if it provides the crucial information required for succeeding in a given game activity. It may be more critical for BVI gamers while not as critical for the seeing gamers.

One potential solution is to enable configuration of HRTFs through an onboarding process prior to experiencing a given VR game. This process may entail active user engagement by incorporating aspects of the given game into the assessment of which set of HRTFs may be superior. This process draws from the ideas from prior studies on HRTF selection methods [17, 18, 19, 20, 21, 22, 23]. These methods rely on objective localisation metrics or subjective preference (i.e., the users own choice) during static localisation tasks where the sound being localised is presented as short bursts. A different approach is to provide a continuous sound stimulus in a localisation task, thereby enabling the user to use active dynamic listening cues which may better resemble an actual VR experience. In active dynamic localisation tasks the HRTFs change with head movement, hence the task typically culminates in an assessment of the frontal HRTFs when the user has located a sound source.

In the present study, we have developed an experimental VR platform utilising a classical 'shooting down sound sources' task with the purpose of exploring which tracking characteristics predict the best performance HRTFs and to assess search patterns during the dynamic localisation task (the onboarding process). The experimental platform will be evaluated across different sound stimuli and HRTFs. We hypothesise that if the non-individual HRTFs are a 'good' match for a given user, then the search pattern will reveal fewer localisation errors during dynamic search. Conversely, if the HRTFs are a 'weak' match then the search pattern may settle or even end in directions where the cues are ambiguous. This study reports data from experiments run on the platform with a normal seeing population exposed to different sound stimuli and HRTFs.

2 Methods

An experimental test platform inspired by the classical 'shooting down sound sources' experiment [24] has been developed using a standard VR platform. The test platform renders a virtual environment with both visible and audible sound sources.

2.1 Test platform

The test platform was developed in the Unity engine [25] for the Meta Quest Pro VR head-mounted display (HMD) and two Meta Touch Pro controllers. The platform ran on a PC connected wirelessly to the HMD through a dedicated router, enabling the experiment operator to monitor the virtual environment (VE).

The VE was a large empty sandbox environment with a standard skybox and horizon, and a large blue plane illustrated the floor. Controllers, hands, arms, torso and the lower body were graphically rendered according to the participant's real positioning using inside-out body tracking. The VE had anechoic acoustic conditions and the sound source had omni-directional directivity with distance falloff and did not generate any reflections. The sound was rendered over a pair of Beyerdynamic DT-990 open-back headphones placed over and connected directly to the HMD. The experiments was run in a large 7 x 8 x 3 meter sound-treated room.

2.2 Experiment structure

The experiment was run in two rounds on separate days for each participant, and each round exposed two different sound stimuli, for a total of four stimuli. Each round comprised four sessions separated by breaks with a duration equal to the session, each session comprised five blocks which were completed in five to seven minutes. Thus, the two sound signals were presented twice per round and the presentation order was balanced to avoid order effects. In each session, the first block was always a familiarisation block with 15 trials where the sound sources are visible, where each trial represented a 'shooting' task. The following three blocks contained 15 trials with invisible sound sources with the sound rendered with different HRTFs. The order of the HRTFs used in the three blocks were balanced using a Latin Square. The final block repeated the same HRTF used in the second block, to assess whether training effects had occurred or the participant was becoming fatigued.

2.3 'Shooting' task

Participants were instructed to search for the sound and then 'shoot' it down with either of the two controllers by pulling a trigger. Feedback was provided after 'shooting' in the invisible trials by briefly displaying the actual sound source location. After a trial ended, the sound was relocated to a new position and a new trial began immediately. White laser beams were projected outwards from the virtual controllers to convey the aiming direction of each controller.

The participants were standing in the center of the room and were provided the information that the sound sources could occur in any random direction and they were instructed to explore the space by moving around and moving their head. The participants were free to move within an unobstructed area, approximately 3 x 3 meters, which was defined by a safety zone on the HMD. Furthermore, participants were instructed to point the controller towards the direction they hear the sound when they attempt to 'shoot'. No further instructions were given on how to localise or aim at the sound sources. Each round of the experiment was estimated to last approximately one hour, wherein approximately half of the time was spent on actively engaging in the 'shooting' task.

2.4 Spatialised sound

The sound was spatialised using Steam Audio SDK for Unity (version 4.7.0) [26] using non-individual head-related impulse responses (HRIRs) measured on human subjects from 193 directions [27]. DC-offset was removed during pre-processing of the HRIRs and conversion to SOFA (Spatially Oriented Format for Acoustics). Three unique HRIR datasets were used in the invisible trials and a different dataset was picked for the familiarisation block, so as to avoid contributing a training effect.

Fifteen sound source locations were randomly generated for every single block in each session and are located on a sphere with a 2.3 meter radius to the approximate center of the participant's head ($x, y, z = (0, 0, 1.7)$) and within a $\pm 45^\circ$ elevation band. The sound source locations were generated using a stratified, staggered sampling scheme designed to approximate a uniform coverage while avoiding a strictly regular grid and clustering of nearby locations. Two space coverage (SC) metrics [16] were computed for the sampled grid: $sc_{angle} = 48^\circ \pm 14.8^\circ$, representing the density of the positions, and $sc_{shape} = 0.83^\circ \pm 0.07^\circ$, representing the distribution's homogeneity. The two metrics reflects a moderately homogeneous distribution, given the gaps at the polar hemispheres inflating the sc_{angle} and standard deviation.

Four different types of sound stimuli (see duration and waveforms in Fig. 1) were used in the eight sessions, so that each stimuli was repeated once with a different HRTF order. All sound stimuli were scaled to a predefined root mean square level and participants were instructed to adjust the sound level to their preferred level during the familiarisation block. The four stimuli were white noise; recording of Danish speech of a fairytale reading by a female speaker [28]; recording of a burning fire [29]; and two sonar beeps [30]. The three latter stimuli were chosen due to their frequency content (see magnitude spectra in Fig. 1), transient content and especially because the sound stimuli are relevant to different game experiences. The speech and fire stimuli have relatively broad spectral content, while the sonar beep has a single peak right below 1000 Hz and a ringing-out effect.

2.5 Data collection

Before starting the task, a baseline Simulator Sickness Questionnaire (SSQ) [31] (translated to Danish and verified in [32]) was answered during the introduction and again after each session. Comparison of SSQ symptom ratings per participant indicated no changes throughout the experiment rounds. Session, block and trial handling along with tracking of position and orientation of the HMD, controllers, torso and virtual character, and logging of discrete task events (e.g., trigger pulls, angular errors, response time, source location) were handled by a custom script utilising Unity Experimental Framework [33].

2.6 Subjects

Fifteen participants (four women and eleven men) completed the first round of the experiment, thereof five participants (all men) also completed the second round (which comprised sonar beep and fire stimulus) on a separate day. The participants' age ranged from 24 to 29 (mean age=25.5 and SD=1.5). The fifteen participants had various prior experience with VR, ranging from no experience to weekly usage.

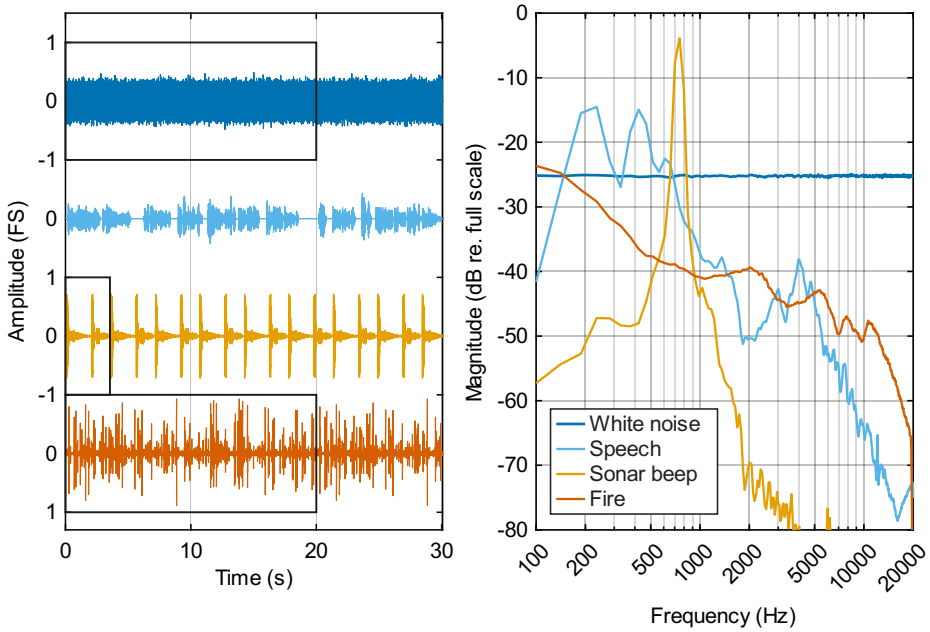


Fig. 1: Waveforms (left panel) and magnitude spectrum (right panel) of the four sound stimuli. The boxes around waveforms indicate length of each stimulus before waveform repetition. The speech recording is cropped to 30 seconds. Full scale level is represented by ± 1 and 0 dB, respectively.

2.7 Data analysis

The angular error (angular difference between 'target' direction and the 'shot' direction indicated by the HMD and controller) is logged for every trial. Summary statistics were analysed for angular distance metrics and response time. Angular distance travelled and angular speed of participants' movement, and trajectories for controller and HMD pointing direction onto a unit sphere have been computed based on the controller and HMD tracking data.

Trajectory clustering analyses were performed on HMD tracking data. For each trial, raw trajectories were first transformed into a 'target'-centred reference frame, so as to align all sound source directions to a single forward axis, and then resampled to a fixed length using spherical linear interpolation [34], while preserving motion on the unit sphere. Dimensionality reduction was then performed using principal component analysis (PCA) and the first ten principal components were retained for clustering. Clustering was conducted using (k)-medoids [35] in the PCA space, evaluated over k=1 to 4 and the (k)-solution was repeated with 50 random initialisations and the solution minimising overall within-cluster distance was retained. An archetype representative for each cluster was defined as medoid trials (observed trajectories).

3 Results

The participants completed a total of 6000 trials, wherein 4800 trials were invisible. Only two single trial responses were discarded due to being registered below a threshold of 500ms response time (early discharge). The mean response time across all trials was 4.5 seconds (standard deviation (SD)=2.2 seconds) and varies between approximately one second for the shortest and 18 seconds for the longest trial. White noise and fire stimuli were on average the fastest trials at approximately 3.6 seconds, while speech and sonar beep stimuli were on average 5.2 seconds.

3.1 Angular distance metrics and elapsed time

Analysis of angular distance metrics indicate that active head movement nearly eliminates front-back reversals entirely (0.1% of trials). Lateral and polar errors are shown in Fig. 2. White noise and fire stimuli have the lowest polar errors across the three HRTF conditions,

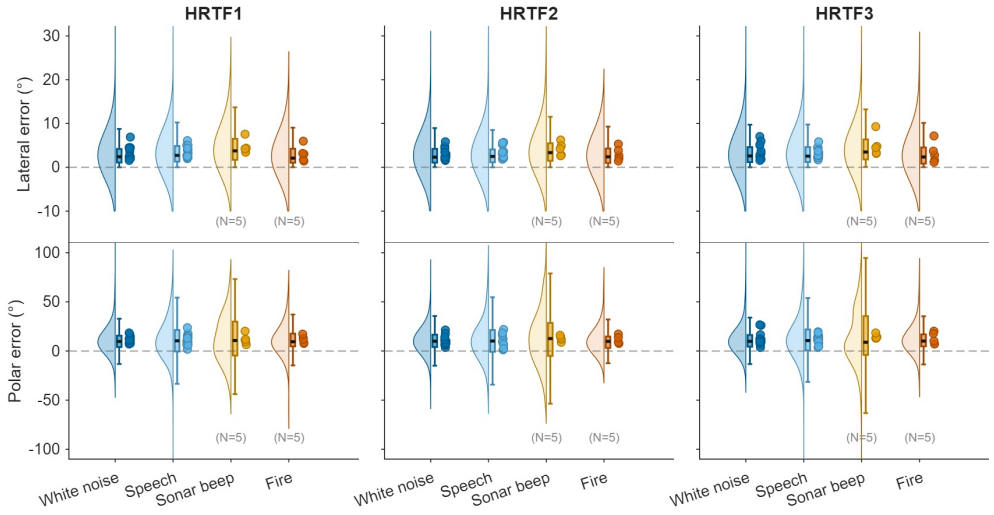


Fig. 2: Raincloud plots of lateral error (first row), polar error (second row), and each HRTF (each column) for the four sound stimuli computed from the controller pointing direction and the 'target' direction. Jittered data points shows per subject x session mean errors and the box plot shows the median (black line) and quartiles.

likely due to having a constant non-transient response and the broadest frequency content. From the angular distance analysis, we find that none of the participants benefitted from a particular HRTF out of the three. The mean across all invisible trials is 14.5° (SD=14.6°). Likewise, for the HMD, the mean is 14.4° (SD=13.2°). Hence, HMD pointing is near-consistently as accurate as controller pointing, despite the participants only being instructed to point towards and 'shoot' the sound sources with the hand-held controller while not being directly instructed to point their head towards the sound sources at any given moment.

3.2 'Shooting'-trial search pattern

Fig. 3 illustrates the angular error between the 'target' source direction and the pointing directions that are determined from the hand-held controller and the HMD as a function of time. The angular speed of the trackers is also determined and plotted as a function of time. During the initial four seconds of the trial the participant remains somewhat still, hence the angular error is relatively unchanged for both trackers and the angular speed is low. The search process starts around five seconds, where the speed increases and the angular error decreases slowly until around ten seconds, where the HMD's angular error becomes very low. Then, the tracker speeds pick up for a moment around 12.5 seconds while the angular error increases, then decreases and eventually the 'shoot' is made at 15.7 seconds. When the 'shot' is fired the controller reaches a very low angular error, while the HMD is slightly further away from the target'. This search pattern indicates that the participant may initially have located the sound at around ten seconds, but then perhaps needed some assurance before confirming the 'shot' with the controller.

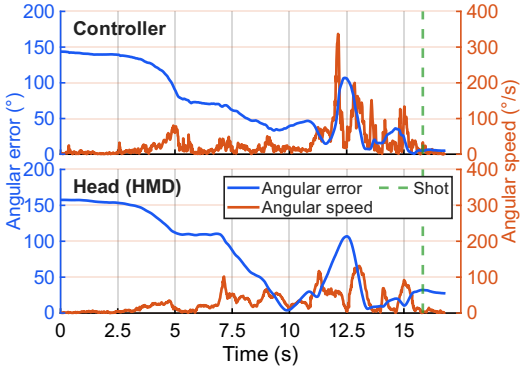


Fig. 3: Angular errors and angular speed versus time for trajectory mapped on a unit sphere of the hand-held controller (top panel) and the head (bottom panel) for one 'shooting' trial. Vertical dashed lines indicate the time of 'shooting'.

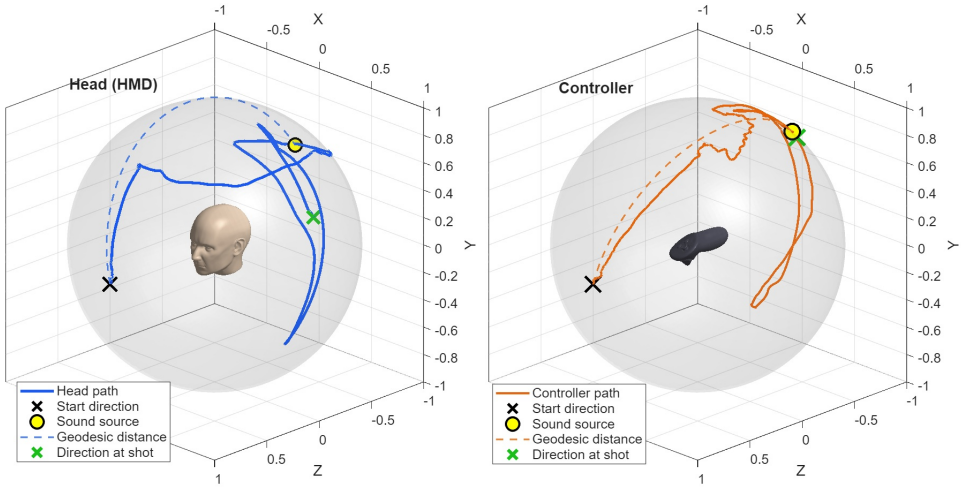


Fig. 4: Trajectory for head-mounted display and controller pointing onto a unit sphere from start position (black X) to 'shot' position (green X) for a single 'shooting' trial. Circle indicates the 'target' sound source and the dashed line indicates shortest distance on sphere from start direction to 'target'. The head and controller indicate the respective start pointing directions.

Fig. 4 reveals the tracked path of the head (HMD) and the controller for the same search pattern in Fig. 3. The start direction is located at a hemisphere opposite to the 'target' location, and the paths move close to the 'target' along the search period before taking a 'detour' downwards into an adjacent hemisphere before returning to the 'target' area. The detour remains on a cone-of-confusion equal to or nearby the target. At the time the 'shot' is made the controller path is close to the 'target' direction (approximately 5° angular error), while the HMD is further away (32° angular error). The duration of the search and the necessity to search into a different hemisphere during the search, perhaps for confirmation, does not suggest that the localisation cues are strong enough, despite the fact the controller pointing managed to locate the 'target' with good precision.

Fig. 5 and 6 illustrate another search pattern for the two trackers, where the initial start direction is in another hemisphere than the 'target' direction. Fig. 5 reveals that within the first two seconds of the search, both angular errors slightly increase and decrease (a reversal) before increasing to a maximum at 1.5-1.8 seconds. Fig. 6 shows that the trajectory moves in an opposite direction compared to the shortest path from start to 'target'. This search pattern may be due to weak localisation cues which make the participant uncertain of which direction to turn towards. At about 8.5 seconds a large spike in controller movement is seen, this is likely the participant attempting to align the controller pointing direction with the participant's head direction before making the 'shot'. However, as is seen in the plotted angular errors this attempt failed and the controller's movement overshoots greatly and the angular error becomes larger than for the HMD. In this

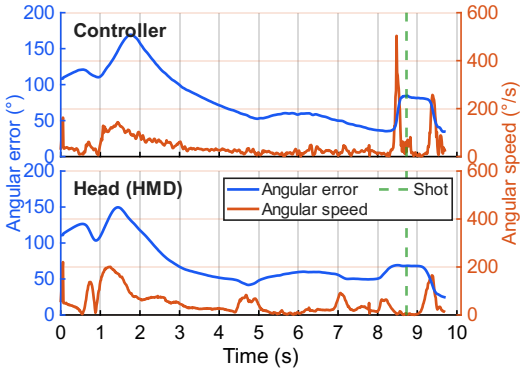


Fig. 5: Angular errors and angular speed versus time for trajectory mapped on a unit sphere of the hand-held controller (top panel) and the head (bottom panel) for one 'shooting' trial. Vertical dashed lines indicate the time of 'shooting'.

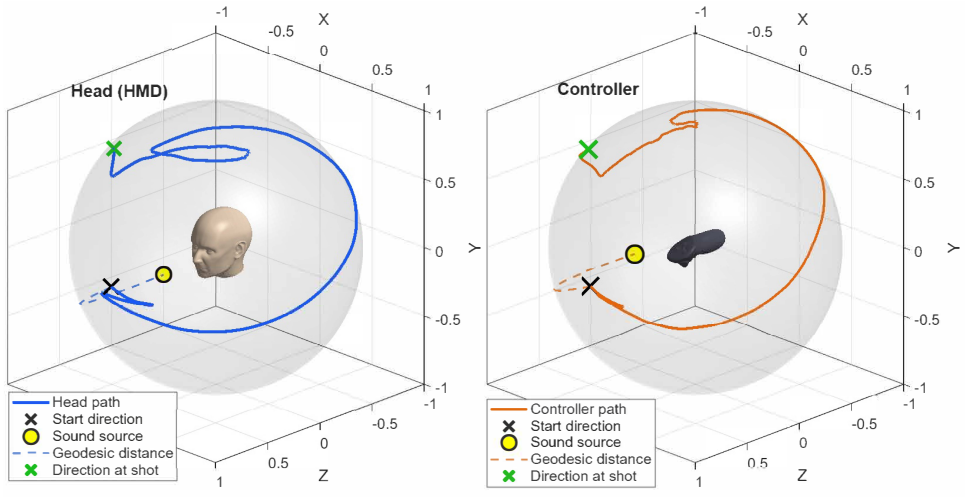


Fig. 6: Trajectory for head-mounted display and controller pointing onto a unit sphere from start position (black X) to 'shot' position (green X) for a single 'shooting' trial. Circle indicates the 'target' sound source and the dashed line indicates shortest distance on sphere from start direction to 'target'. The head and controller indicate the respective start pointing directions.

trial, the azimuth offset is somewhat equal for both trackers' final direction (-3.0° and -5.8°), whereas the controllers elevation offset is much higher (83.9°) than the HMD (68.3°). Hence, in the end the localisation cues were only strong enough for the participant to correctly identify the azimuth direction of the 'target' after overcoming the initial misguided search direction.

3.3 Prototype search patterns

From the collected data, we have identified various search patterns that may indicate weak localisation cues, these are: the direction of initial search is opposite to the 'target' direction or the shortest distance (great-circle) to the 'target'; the search pattern contains several search direction reversals; the search pattern greatly overshoots the 'target' direction or dwells on a cone-of-confusion different from the 'target'; the 'shot' direction is in a different hemisphere than the 'target'; a search pattern with a very long elapsed time; dwelling near targets' for a time period before deciding to 'shoot'; and a search pattern which takes an 'excursion' or 'detour' e.g., into a different hemisphere. Several search patterns have been identified on trials that seemingly end with a relatively low angular localisation error, as seen in Fig. 3 and 4.

Prototype search patterns have also been identified using PCA and (k)-medoid clustering. Clustering was performed on the full dataset and on a restricted dataset of long-latency trials (response time >3.5 seconds), to characterise high-uncertainty localisation rather than population-wide trajectory patterns. Robustness was assessed by analyses of alternative thresholds and individual stimulus datasets.

Four clusters for white noise and speech trials (n=2609) are shown in the left panel of Fig. 7. Cluster 1, 2 and 4 display archetypal patterns with a near-direct trajectory towards the correct cone-of-confusion relative to the 'target' followed by small elevation search motion before providing a response near the 'target'. Cluster 3 displays an archetypal pattern that contain several direction reversals, both polar and lateral searching detours, and an in-accurate elevated response. Four clusters for white noise and speech trials with a response time >3.5 seconds (n=1638) are shown in the right panel of Fig. 7. Clusters 1, 2 and 3 display archetypal patterns similar to the prior cluster 1, 2 and 4 for the full dataset, where the final response direction is nearby the 'target' direction, except for cluster 2. Cluster 3 displays an archetypal pattern which contains a large initial lateral search strategy followed by a polar search after locating the correct cone-of-confusion associated with the target

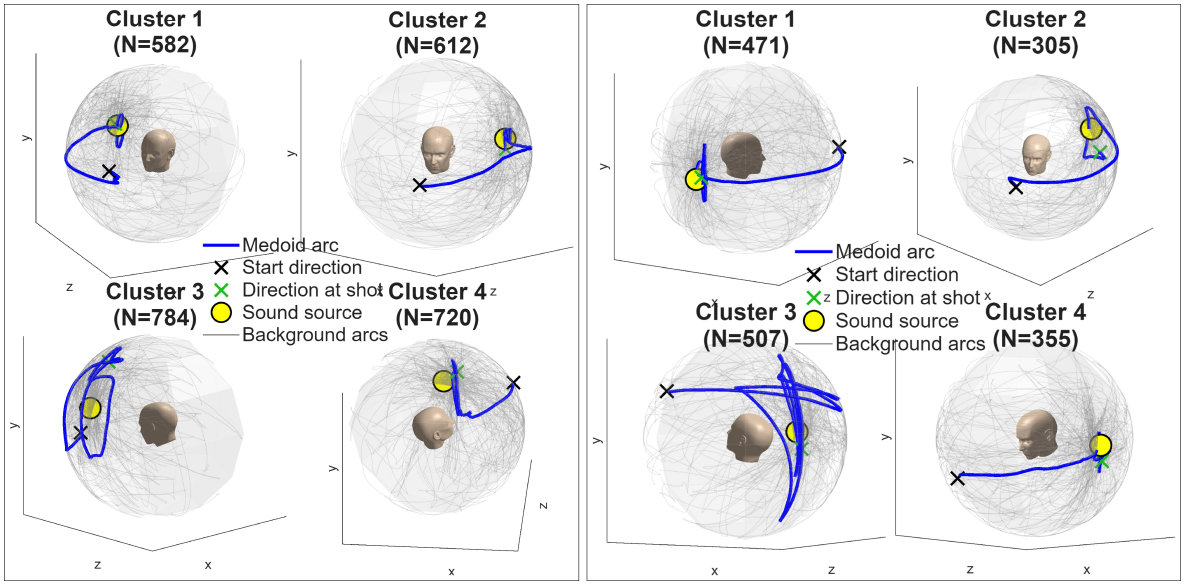


Fig. 7: Trajectory clusters for white noise and speech trials for clusters across entire dataset (left panel) and clusters across dataset with response time >3.6 seconds (right panel). Blue trajectory indicates the medoid trial for each cluster, the head points towards the start direction (black X), green X indicates direction at shot, and the yellow circle indicates target direction. 50 trajectory arcs are plotted in the background.

direction.

4 Discussion

The present study employed multiple HRTFs and sound stimuli in the localisation task, which the participant is largely unfamiliar with at the beginning. This might not be the best-case scenario for successful localisation, and determining the best HRTFs. On the other hand, the stimuli had broad spectral content, allowing the participant to use the salient features of the HRTFs for localisation. Using game-relevant sound stimuli in an onboarding process provides the added benefit of familiarisation before game start.

Preliminary results indicate that on average, the participants were able to localise the sound sources with a decent angular error (approximately 14.5°), within a relatively short elapsed time of 4.5 seconds, and with near zero front-back confusions. However, during our analyses we found various search patterns that convey a strong lack of localisation cues, even in trials where the final angular errors indicated correct localisation of the sound source. Further analysis should quantify the importance of the various trajectory metrics (e.g., path length, speed, direction and time spent searching in different confusion cones) and compare these search patterns and their implication to existing localisation metrics and in-particular subjective perception.

5 Conclusion

An experimental platform for testing a user's localisation abilities in a simplistic virtual environment has been implemented. A classical 'shooting down sound sources' task is suggested as a test paradigm for a game onboarding process to introduce the user to gameplay mechanics and monitor localisation ability. The sound sources' location in space are computed on the fly during onboarding to the game and game-relevant sound stimuli is rendered binaurally to the user. Search patterns are identified and analysed to determine weak localisation cues and possible ambiguities during the localisation task. The motivation is to enable the determination of the most favourable HRTFs for the individual user, a feature yet to be implemented and further explored. Further studies may explore whether it is possible to find the objectively 'best'-matching HRTFs during game onboarding, and user satisfaction may also be assessed for the suggested and alternative paradigms, which are steered by user preferences rather than objective assessment.

6 Acknowledgement

We want to thank our participants as well as our collaborators Antonia Line Krummheuer, Thomas Bjørner, Stine Froulund Astorp, and Nicoleta Radu. Technical assistance by Claus Vestergaard Skipper is also greatly appreciated. This study was funded by the Independent Research Fund Denmark DFF2 grant 4264-00072B.

References

[1] Morgan, M., "Sensory perception: Supernormal hearing in the blind?" *Curr. Biol.*, 9(2), pp. R53–R54, 1999, doi:10.1016/S0960-9822(99)80009-8.

[2] Voss, P., Lassonde, M., Gougoux, F., Fortin, M., Guillemot, J.-P., and Lepore, F., "Early- and Late-Onset Blind Individuals Show Supra-Normal Auditory Abilities in Far-Space," *Curr. Biol.*, 14(19), pp. 1734–1738, 2004, doi:10.1016/j.cub.2004.09.051.

[3] Schepers, I. M., Hipp, J. F., Schneider, T. R., Röder, B., and Engel, A. K., "Functionally specific oscillatory activity correlates between visual and auditory cortex in the blind," *Brain*, 135(3), pp. 922–934, 2012, doi:10.1093/brain/aws014.

[4] Boroujeni, F. M., Heidari, F., Rouzbahani, M., and Kamali, M., "Comparison of auditory stream segregation in sighted and early blind individuals," *Neurosci. Lett.*, 638, pp. 218–221, 2017, doi:10.1016/j.neulet.2016.12.022.

[5] Shim, H. J., Go, G., Lee, H., Choi, S. W., and Won, J. H., "Influence of Visual Deprivation on Auditory Spectral Resolution, Temporal Resolution, and Speech Perception," *Front. Neurosci.*, 13, 2019, doi:10.3389/fnins.2019.01200.

[6] Battal, C., Occelli, V., Bertonati, G., Falagiarda, F., and Collignon, O., "General Enhancement of Spatial Hearing in Congenitally Blind People," *Psychol. Sci.*, 31(9), pp. 1129–1139, 2020, doi:10.1177/0956797620935584.

[7] Sabourin, C. J., Merrikhi, Y., and Lomber, S. G., "Do blind people hear better?" *Trends Cogn. Sci.*, 26(11), pp. 999–1012, 2022, doi:10.1016/j.tics.2022.08.016.

[8] Fine, I. and Park, W. J., "Do you hear what I see? How do early blind individuals experience object motion?" *Philos. Trans. R. Soc. B Biol. Sci.*, 378(1869), 2023, doi:10.1098/rstb.2021.0460.

[9] Zwiers, M. P., Opstal, A. J. V., and Cruysberg, J. R. M., "A Spatial Hearing Deficit in Early-Blind Humans," *J. Neurosci.*, 21(9), 2001, doi:10.1523/JNEUROSCI.21-09-j0002.2001.

[10] King, A. J., "What happens to your hearing if you are born blind?" *Brain*, 137(1), pp. 6–8, 2014, doi:10.1093/brain/awt346.

[11] Moller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D., "Binaural Technique: Do We Need Individual Recordings?" *J. Audio Eng. Soc.*, 44(6), pp. 451–469, 1996.

[12] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L., "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Am.*, 94(1), pp. 111–123, 1993, doi:10.1121/1.407089.

[13] Fels, J., Braren, H., and Bomhardt, R., "Comparison of the front-back confusion rate of individual and individualized head-related transfer functions," in *DAGA 2018*, Deutsche Gesellschaft für Akustik e.V. (DEGA), München, 2018.

[14] Cuevas-Rodriguez, M., Gonzalez-Toledo, D., Reyes-Lecuona, A., and Picinali, L., "Impact of non-individualised head related transfer functions on speech-in-noise performances within a synthesised virtual environment," *J. Acoust. Soc. Am.*, 149(4), pp. 2573–2586, 2021, doi:10.1121/10.0004220.

[15] Møller, H., Jensen, C. B., Hammershøi, D., and Sørensen, M. F., "Using a typical human subject for binaural recording," in *In Proc. of 100th Audio Engineering Society Convention*, 4157, pp. 1–18, Audio Engineering Society, 1996.

[16] Poirier-Quinot, D., S. Lawless, M., Stitt, P., and F.G. Katz, B., "HRTF Performance Evaluation: Methodology and Metrics for Localisation Accuracy and Learning Assessment," in *Advances in Fundamental and Applied Research on Spatial Audio*, IntechOpen, 2022, doi:10.5772/intechopen.104931.

[17] Poirier-Quinot, D. and Katz, B. F., "On the improvement of accommodation to non-individual HRTFs via VR active learning and inclusion of a 3D room response," *Acta Acustica*, 5, p. 25, 2021, doi:10.1051/aacus/2021019.

[18] Poirier-Quinot, D., Lawless, M., and Katz, B., "When HRTF training doesn't work," in *Proc. of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, pp. 2339–2346, European Acoustics Association, 2024, doi:10.61782/fa.2023.0904.

[19] Seeber, B. U. and Fastl, H., "Subjective Selection of Non-individual Head-related Transfer Functions," in *Proc. of the 2003 Int. Conf. on Auditory Display*, pp. 259–262, International Community For Auditory Display, Boston, MA, USA, 2003.

[20] Katz, B. F. G. and Parseihian, G., "Perceptually based head-related transfer function database optimization," *J. Acoust. Soc. Am.*, 131(2), pp. 99–105, 2012, doi:10.1121/1.3672641.

[21] Daugintis, R., Barumerli, R., Picinali, L., and Geronazzo, M., "Classifying Non-Individual Head-Related Transfer Functions with A Computational Auditory Model: Calibration And Metrics," in *ICASSP 2023 - 2023 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, IEEE, 2023, doi:10.1109/ICASSP49357.2023.10095152.

[22] Bahu, H., "Auditory localization in the context of non-individual binaural synthesis," in *Summary of the PhD*, pp. 1–9, Institute of Research and Coordination in Acoustics and Music (IRCAM), 2016.

[23] Zagala, F., Noisternig, M., and Katz, B. F. G., "Comparison of direct and indirect perceptual head-related transfer function selection methods," *J. Acoust. Soc. Am.*, 147(5), pp. 3376–3389, 2020, doi:10.1121/10.0001183.

[24] Sandvad, J., "Dynamic Aspects of Auditory Virtual Environments," in *Proc. of 100th Audio Engineering Society Convention, Copenhagen, May 11-14, 1996*, pp. 1–14, Audio Engineering Society, Copenhagen, 1996.

[25] "Unity Engine," https://unity.com/, 2025.

[26] "Steam Audio SDK," https://valvesoftware.github.io/steam-audio/, 2025.

[27] Møller, H., Friis Sørensen, M., Hammershøi, D., and Boje Jensen, C., "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, 43(5), pp. 300–321, 1995.

[28] Marryat, F., "Børnene i Nyskoven," https://archive.org/details/bornene_nyskoven_1212_librivox, 2012, read in Danish by Lulularsen.

[29] SoundReality, "Fire Forest 2," https://pixabay.com/sound-effects/nature-fire-forest-2-377740/, 2025.

[30] MrLindstrom (Freesound), "SonarPing," https://pixabay.com/sound-effects/film-special-effects-sonarping-38269/, 2022.

[31] Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G., "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness," *Int. J. Aviat. Psychol.*, 3(3), pp. 203–220, 1993, doi:10.1207/s15327108ijap0303_3.

[32] Eriksen, A. M., Jelic, G., Dalsgaard, G. T., Pedersen, L. S., Østergaard, M., and Damborg, S., "Cyber Sickness in Virtual Reality," 4th semester project, product and design psychology (group 480), Aalborg University, Department of Electronic Systems, 2020.

[33] Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., and Mushtaq, F., "Studying human behavior with virtual reality: The Unity Experiment Framework," *Behav. Res. Methods*, 52(2), pp. 455–463, 2020, doi:10.3758/s13428-019-01242-0.

[34] Shoemake, K., "Animating rotation with quaternion curves," *SIGGRAPH Comput. Graph.*, 19(3), p. 245–254, 1985, doi:10.1145/325165.325242.

[35] Kaufman, L. and Rousseeuw, P. J., *Finding Groups in Data: An Introduction to Cluster Analysis.*, John Wiley, 1990, doi:10.1002/9780470316801.