USER HRTF SELECTION FOR 3D AUDITORY MIXED REALITY

Rishi SHUKLA[1], **Rebecca STEWART**[2], and **Mark SANDLER**[1]

[1]*Centre for Digital Music*, **Queen Mary University of London**, London, UK
[2]*Dyson School of Design Engineering* , **Imperial College London**, London, UK

ABSTRACT

We introduce a novel approach for personalisation of an efficient 3D binaural rendering system designed for mobile, *auditory mixed reality* use cases. A head-related transfer function (HRTF) ranking method is outlined for users of real-time, interactive sound and music applications. Twenty participants tested the approach and its impact on their capacity to locate a continuous musical sound rendered in varying 3D positions. Analysis of HRTF rankings across three separate sessions reveal encouraging levels of reliability amongst some participants. Patterns of interaction show a significant benefit to horizontal precision that results from the selection process. In contrast, length of system exposure (rather than HRTF preference) demonstrates a significant degree of improvement to aspects of vertical perception and overall speed of response, with no detriment to horizontal accuracy. These findings provide an initial basis from which to consider priorities in the design of audio-only immersive applications and accompanying methods for effective user controlled personalisation.

1. INTRODUCTION

Since the 1990s, sound computing research has explored spatially located interactive audio within real world contexts. The principles and applications of binaural synthesis (i.e. 360° sound simulation over headphones) were given prominent discussion in Begault's major work on 3D audio for multimedia [1]. Bederson's virtual tour guide presented an early design for personalised real-time audio intervention to augment physical environments [2]. Possibilities for blending acoustic and digitally situated sound were later investigated with closed-ear headphones (Cohen's experiment with artificially spatialised sound sources [3]) and wearable speakers (Sawnhey's Nomadic Radio [4]).

Mobile computing has brought renewed focus on personalised interactive sound design that supplants or supplements environmental acoustics, for example through multilayered spatial auditory display to encourage exploration of art exhibitions [5]. 'Hearable' headsets have further accelerated development. These devices feature integrated orientation and often positional sensing to enable immersive audio spatialisation. Direct or electronically assisted openness to environmental sound is also enabled in some instances. Recent evaluation of these devices has highlighted potential affordances and some initial interaction design recommendations. It has also introduced new working definitions to distinguish the variety of hardware features and modes of experience that coexist in *auditory mixed reality* (AMR), a term that 'encapsulates any auditory VR and AR experiences' (i.e. forms of sonic virtual or augmented reality) [6]. These technologies have recently prompted structured experimentation with creative social gameplay [7], as well as direct industry research addressing how audio virtual reality might be applied to music discovery activity [8].

Research into personalisation of spatial rendering for these audio-only contexts is limited. The potential benefits to user orientation and sense of immersion from incorporating HRTF personalisation have been highlighted in previous spatial audio experience studies as an area requiring structured investigation [5, 9]. This paper presents and evaluates a method for user selection of preferred HRTF sets via a custom binaural rendering system for mobile, interactive spatial audio. Section 2 discusses the limitations of prior work on HRTF selection for end-user audio-only contexts and specific considerations therein. Section 3 outlines the method conceived for 3D HRTF comparison and the experimental protocol for measuring its performance. Results of the evaluation are presented in section 4, followed by discussion of outcomes and summary conclusions in sections 5 and 6.

2. BACKGROUND

HRTF sets are location dependent filters that simulate spatialisation of sources around a listener. HRTF sets comprise multiple head-related impulse responses (HRIRs), which are measured within the left and right ear of a human or dummy head, using an excitation source placed at incremental surrounding positions. The efficacy of HRTF processing is dependent on both the density of the measurements, and correlation between the HRIRs and features of the listener's morphology that affect spatial perception (chiefly the shape of their head, pinnae and upper body) [1, 10]. Use of a generic or poorly matched HRTF set is liable to impact sound localisation accuracy, discrimination between sources in front/behind and above/below, externalisation (i.e. sense that a sound is emanating from outside the lister's head), and tonal clarity.

A growing body of research demonstrates the efficacy of parametric methods for either selecting or simulating best-fitting HRTF sets according to user morphology [11–14]. However, precise and reliable acquisition of anthropometric head, ear and torso features in the case of mobile system end-users is a nontrivial challenge. Selection of a best-fitting HRTF set from a database of alternatives is another established strategy for customising binaural rendering [15, 16]. However, the latter approach – which is the area of concern for the remainder of this paper – also raises specific design problems for the context of AMR.

2.1 Working criteria for end user HRTF selection

The authors have previously discussed the difficulties of applying established HRTF evaluation methodologies within an end user system [17]. Objective approaches require subjects to repeatedly identify perceived locations of sources placed in various spatial positions, for alternate HRTF sets. This is by its nature time-consuming and typically relies on a laboratory-style format and rapid serial responses to short test signal stimuli [18–20]. Subjective approaches, by contrast, ask users to rate the effect of alternate HRTF sets using separate criteria, such as consistency in motion/trajectory, hemispherical distinction or degree of externalisation [21, 22]. Evidence suggests that repeatability of such qualitative judgements is contingent on listener expertise [23–26].

A design for HRTF selection in 2D was previously devised using interactive holistic A/B comparison, with recorded music as the stimulus signal. The approach was evaluated against these four criteria with encouraging outcomes [17].

- *Reliability* – clear and consistent selection outcomes
- *Validity* – spatial fidelity that is fit for purpose
- *Usability* – of potential benefit to any end user
- *Efficiency* – a duration acceptable for the use case

2.2 Considerations for 3D audio-only mobile contexts

Adapting the 2D holistic approach for 3D judgement of HRTFs in mobile use cases presents three challenges:

1. Thorough comparative judgement of 3D involves a much wider range of spatial positions and trajectories, all of which must be adequately explored to make valid selections.
2. Mobile and 'hearable' devices have either restricted or no visual display, meaning that judgements should be made without reference to dynamic graphics or complex interfaces that would be necessary for audiovisual interaction.
3. A clear indication of selection repeatability is necessary in light of the above complexities.

This research deploys a virtual vector base amplitude panning (VBAP) [27] system developed on an open-source embedded Linux platform. The approach uses eight HRIR pairs to simulate a sparse, pseudo-spherical virtual loudspeaker array. Individual sound sources are positioned via VBAP processing prior to binaural encoding. Five virtual speakers are located on the horizontal plane (0° elevation) at 0°, -60°, 60°, 120° and -120°. Three further speakers are placed around the median plane (0° azimuth) at 90° (directly above), -45° (below front) and -135° (below back). VBAP has been shown to render with favourable levels of spatial and tonal consistency in this virtual configuration [28] and more generically [29]. However, specific limitations to vertical cue representation in a sparse array layout, such as that defined above, are also well known [30].

3. METHODOLOGY

A selection method was devised and evaluated against the four criteria in section 2.1. Participants undertook three identical study sessions (to assess *reliability*). In each session, part one comprised of the HRTF selection procedure (to test *usability* and *efficiency*) and part two consisted of a follow-up objective localisation task (to measure *validity*).

3.1 Participation

Twenty-one participants (aged 25-45, 6 female and 15 male) were recruited on to the study, which was approved by the Queen Mary University of London (QMUL) Ethics Committee (reference 2038). Each session was approximately 45 minutes, with a minimum of 48 hours between sittings. Participants received £20 in compensation for their time at the end of their third session. All participants were recruited via an open call to staff and doctoral students across QMUL's School of Electronic Engineering and Computer Science. No hearing impairments were declared, other than from two participants who reported occasional and slight tinnitus that neither regarded as pronounced. A questionnaire was used to collect information on musical training, headphone listening habits and prior exposure to binaural audio.

Participants wore a battery powered embedded device running the rendering, study journey and response logging software. They used an iOS device to submit responses and progress through the study. Sennheiser HD650 headphones (without any equalisation for binaural synthesis applied) were secured to participants with an elasticated hairband. The head-tracking sensor was mounted on top of the headphones to counter-rotate the sound scene and take head position readings at 1° angular resolution. Head-tracking was only enabled in part two of the study.

3.2 Part one: HRTF selection

A comprehensive tournament structure used 21 pairwise comparisons between the optimised shortlist of seven human-measured HRTF sets from the LISTEN database [31] identified in [32]. A notable previous investigation into HRTF selection repeatability used stimuli with fixed trajectories that were not responsive to head-tracking [23]. The test pursued here followed a similar approach, but used content derived from recorded music, rather than test tone signals. The comparison stimulus was compiled from excerpts of an anechoic recording of Haydn's Trumpet Concerto in Eb performed on unaccompanied cornet [33].

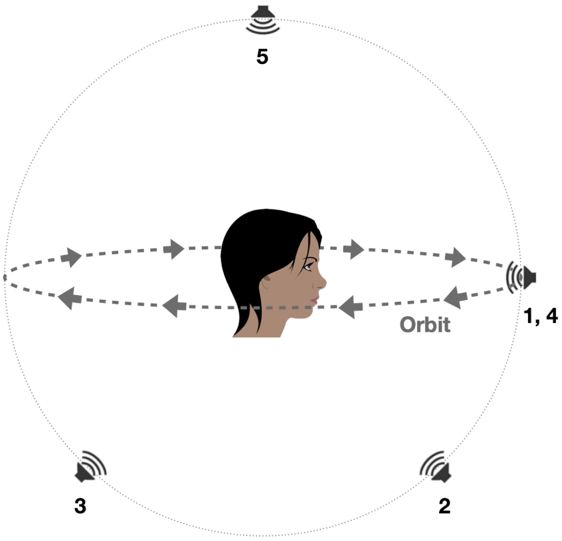


Figure 1. Trajectory for virtual VBAP HRTF comparison

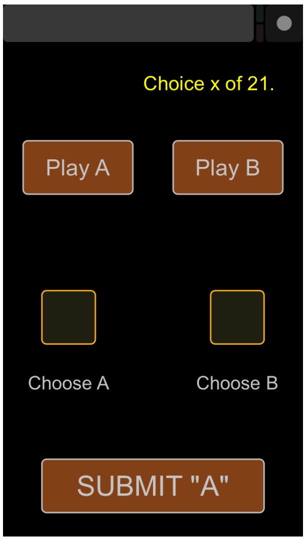


Figure 2. User interface for HRTF preference submission

Figure 3. Localisation testing setup showing equipment worn, personalised calibration point and interaction mode

3.2.1 Trajectory

The trajectory in figure 1 addresses considerations 1 and 2 specified in section 2.2. The horizontal plane orbit consisted of a single sustained note lasting approximately two seconds. Five short bursts in four static positions on the median plane used an even four-note phrase of around one second. The overall stimulus was a little under ten seconds. A key feature of the trajectory was that it passed through all eight virtual speaker locations, which takes advantage of VBAP's amplitude panning approach to spatialisation. Although the trajectory covered just 363 of 64,082 potential coordinates, this small minority focussed on the eight fundamental points from which all locations are rendered (consideration 1). The trajectory was also judged sufficiently short and simple enough to enable purely internalised A/B auditory comparison without reference to dynamic or interactive graphics (consideration 2).

3.2.2 Selection process

Participants used the GUI shown in figure 2 to compare trajectories and submit preferences. They were also given the diagram in figure 1 and an accompanying instruction:

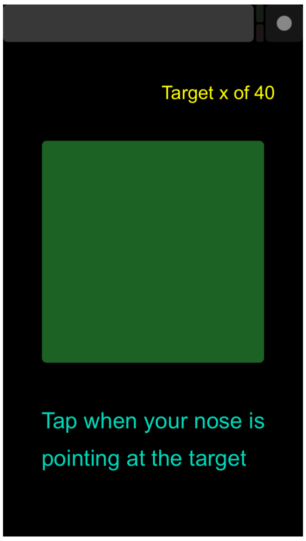
> ***Which has the more convincing 3D effect, excerpt A or B?***
>
> *When comparing A and B, you may wish to consider:*
>
> - ***horizontal accuracy*** *during the orbit and at the four static central positions*
> - *sense of* ***spread*** *between front/back and up/down positions*
> - *sense of* ***distance*** *or how "outside of your head" the sounds seem*

Participants completed one example response to check their understanding of the task before starting. For each response, both the time elapsed and outcome of each comparison was logged automatically. Participants were allowed to listen to either trajectory as many times as they wished, but were forced to listen to A and B at least once in their entirety, before response buttons became enabled.

Both the sequence of comparisons and the order of A/B pairs were randomised for each participant, at all three sittings. For each of their three sessions, the HRTF sets chosen and rejected most often were designated as the preferred and least favoured options. In the event of a draw one of the tied sets was picked at random. These two designations were then used as the best and worst fitting HRTF sets in participants' subsequent localisation test.

3.3 Part 2: Interactive localisation test

The localisation test was conducted following a break of around five minutes. Before starting, the head-tracker was calibrated to a personalised position measured and agreed as approximately directly ahead and level with their eyeline and therefore considered as 0° azimuth, 0° elevation. Figure 3 shows the physical setup of the test environment. The localisation stimulus used 20 seconds of continuous music from the same recording used in part one [33].


Target x of 40

Tap when your nose is pointing at the target

Figure 4. User interface for localisation test submission

3.3.1 *Target locations*

Localiation targets were divided into three strata, so that anticipated shortcomings in upper and lower hemisphere rendering cues could be evaluated independently:

- **at 45° elevation** – seven azimuths of -153°, -102°, -51°, 0°, 51°, 102° and 153°
- **at 0° elevation** – six of the azimuths stated above (0° was not used)
- **at -45° elevation** – the seven azimuths stated above

3.3.2 *Localisation process*

The test used egocentric head-pointing to report perceived source position, comparable to [34]. Participants used the simple GUI in figure 4 with the instruction:

> ***Where is the target sound source?***
>
> *Find the location of the target sound. Point your nose towards what you hear to be the source position.*
>
> *The source will be from somewhere around you and sometimes above or below your ear level. In some cases, you might need to rotate in your seat and/or tilt your head to point accurately.*

Participants completed two example responses to check that they understood what was required before starting the task. For each response, both the time elapsed and variance in head position from target location was logged automatically (as azimuth and elevation co-ordinates). Participants were allowed as much time as they needed to respond for each target. The 20 second excerpt continued on a loop, if necessary, until they registered a response, after which time the next target location began automatically.

Both the sequence of 20 co-ordinates and the order of the two groupings (preferred and least favoured HRTF set) were randomised for each participant, at all three sittings.

Systematic Disagreement			Weak Agreement			Fair or Good Agreement		
C	-0.306	(n)	*A*	0.349	(n)	*B*	0.429	(n)
D	-1.080	(o)	*G*	0.186	(n)	*E*	0.743	(p)
H	-0.095	(o)	*K*	0.075	(n)	*F*	0.437	(n)
J	-0.418	(o)	*O*	0.342	(p)	*I*	0.715	(n)
L	-0.840	(p)	*P*	0.380	(o)	*N*	0.726	(p)
M	-0.795	(n)	*Q*	0.142	(o)	*R*	0.648	(o)
T	-1.151	(p)	*U*	0.258	(o)	*S*	0.510	(o)

Table 1. HRTF selection reliability values and category for each participant (A-T), including binaural experience indicator (n = none; o = occasional; p = practised)

Therefore, a total of 120 data points was recorded for each participant using between a minimum of two (in the event of perfectly repeated best and worst selections) and maximum of six different HRTF sets.

4. RESULTS

All participants completed three study sessions at least 48 hours apart. During the session and on later examination of data, it became evident that one participant had not understood the requirements of the localisation task and had provided responses that did not actively seek out the position of the sound source. This participant's data is reflected in the analysis that follows in part one, but not in part two.

4.1 Part one: HRTF selection outcomes

Between the 21 participants, 63 HRTF selection procedures were completed. The average time taken across these was a mean of 13 and median of 11.8 minutes.

4.1.1 *Ranking method*

For each session, the outcomes of a participant's comparisons were translated into rank order based on the number of times each HRTF set was selected (a maximum of six and minimum of zero occasions). Tied HRTF sets were given a shared ranking at the highest jointly occupied position. So, for example, a ranking list of 1,2,3,4,4,4,7 reflects three HRTF sets gaining a score equal to fourth place.

4.1.2 *Intra-class correlation measurement*

Intra-class correlation (ICC) is a statistical approach used for measuring consistency between different raters to verify the robustness of a rating system [35]. ICC has been used previously to evaluate the reliability of repeated HRTF set ratings expressed by the same raters [26]. The HRTF selection reliability established for each participant via ICC is presented in table 1. Calculation of ICC was achieved using the R statistical computing package, according to the guidance and numerical outcome classifications provided in [35], where: less than 0.0 represents lower than chance levels of agreement (systematic disagreement); between 0.0 and less than 0.4 is an above chance (but weak) level of agreement; from 0.4 to less than 0.6 indicates fair agreement; between 0.6 and less than

0.75 shows good agreement; 0.75 and beyond constitutes excellent agreement.

Details provided by participants about their musical training, headphone listening habits and level of prior exposure to binaural audio were analysed by the groups in table 1 using chi-square and multiple linear regression tests. No relationship was evident between selection consistency and any factor. Each participant's level of experience with binaural audio is shown in table 1 for reference.

4.2 Part two: Interactive localisation outcomes

Three factors mean a reasonable degree of error was to be expected, particularly at the upper and lower strata (45° / -45° elevation). Firstly, even under optimal acoustic conditions, localisation blur of broadband sound immediately in front of a listener is established to be in the order of +/- 3.6° for azimuth and +/- 9° for elevation [36]. Secondly, inaccuracy in head pointing orientation was an further contributor to response error. Bahu et al [34] suggest that, for sources with substantial vertical displacement (57°), head pointed localisation can introduce mean unsigned error of 3° in azimuth and 12° in elevation. Thirdly, sparseness of the virtual speaker array in the upper and lower binaural hemisphere would have degraded spatial representation of sources originating in these areas far beyond optimal acoustic conditions [28,30].

Given these constraints, minimum standards of accuracy were established to evaluate localisation outcomes. For azimuth, a tolerance of +/-15° was used to test whether the rendering system could provide reliable interactive presentation of sound sources at a minimum lateral separation of 30°. For elevation, a +/-22.5° threshold was applied to test simply whether users could reliably distinguish between sources located above, below and on the horizontal plane.

4.2.1 Influence of HRTF selection

Quality of HRTF fit could have impacted both response accuracy and time. Figure 5 shows the distribution of participant outcomes for best and worst HRTFs. Plots show the distribution of participants' overall azimuth and elevation success rate and their mean response duration, for each stratum (45°, 0° and -45°). If there were objective interaction benefits to the HRTF selection procedure, we would expect to see higher successful identification rates and lower mean response times. This is only evidenced clearly in relation to azimuth accuracy in the upper hemisphere (upper and middle plots of column one in figure 5). A Wilcoxon signed rank test confirmed significant improvement in accuracy of azimuth for sources placed on the horizontal plane, when using a preferred HRTF set compared to least preferred ($p = 0.047$). The same non-parametric test for significance did not uncover any other effects from using the best judged HRTF set, for any of the remaining eight metrics in figure 5.

Further analysis was conducted to evaluate the influence of HRTF selection consistency on localisation performance. A Kruskal-Wallis test was conducted between the participant groupings shown in table 1 (but without the participant identified in 4, who was within the 'Weak Agreement' group) and the same nine metrics reflected in figure 5. Significant difference was found in elevation accuracy at 0° ($p = 0.008$, $\chi^2 = 9.575$). Post-hoc Tukey-Kramer analysis identified that the 'Systematic Disagreement' group performed with significantly better accuracy against this metric than the 'Weak' group.

4.2.2 Influence of learning effects

Figure 6 represents the distribution of participant outcomes when viewed as the first and second halves of their sessions (irrespective of best/worst HRTF sequencing, which was always randomised). A Wilcoxon signed rank test confirmed significant improvement in accuracy of elevation identification with sources placed at 45°, for responses given in the second half of localisation trials ($p = 0.041$). The same non-parametric test further identified that responses at 45° ($p = 0.003$) and 0° ($p < 0.001$) were quicker in the second half, without any detrimental impact on azimuth or elevation precision.

5. DISCUSSION

We return to the four criteria put forward in section 2 for evaluating the success of an HRTF selection system.

5.1 Reliability

A third of participants demonstrated a fair to good level of consistency in the rankings that resulted between their three HRTF selection sessions. A further third showed some tendency towards repeating their patterns of selection beyond chance level. The final third returned sets of rankings that actively diverged from each other to a greater degree than chance level.

Although absolute values and proportions of participants between these groups do not indicate a mechanism that could yet be described as reliable, for a significant minority it was possible to attain outcomes that were repeatable to an acceptable level. Given the holistic nature of the comparison judgement (simultaneously considering azimuth, elevation and externalisation) and speed of the overall selection process, the approach shows substantial potential for further development towards more reliable usage. More detailed analysis of the selection process and localisation data presented here will help to identify how the stimulus, trajectory or written guidance outlined in section 3 could be further simplified to focus the comparison process.

5.2 Validity

Analysis showed apparent benefits to azimuth localisation accuracy in the upper hemisphere from preferred HRTF selection, which was significant along the horizontal plane. It is unsurprising that preferred HRTFs were of most benefit across this dimension, where five of the eight virtual speakers reside. Although this finding validates the selection approach in one respect, it is notable that positive elevation detection was increased by general exposure to the localisation task (albeit from a low starting base). This improvement and accompanying increases in response speed occurred independently of best or worst HRTF usage. The

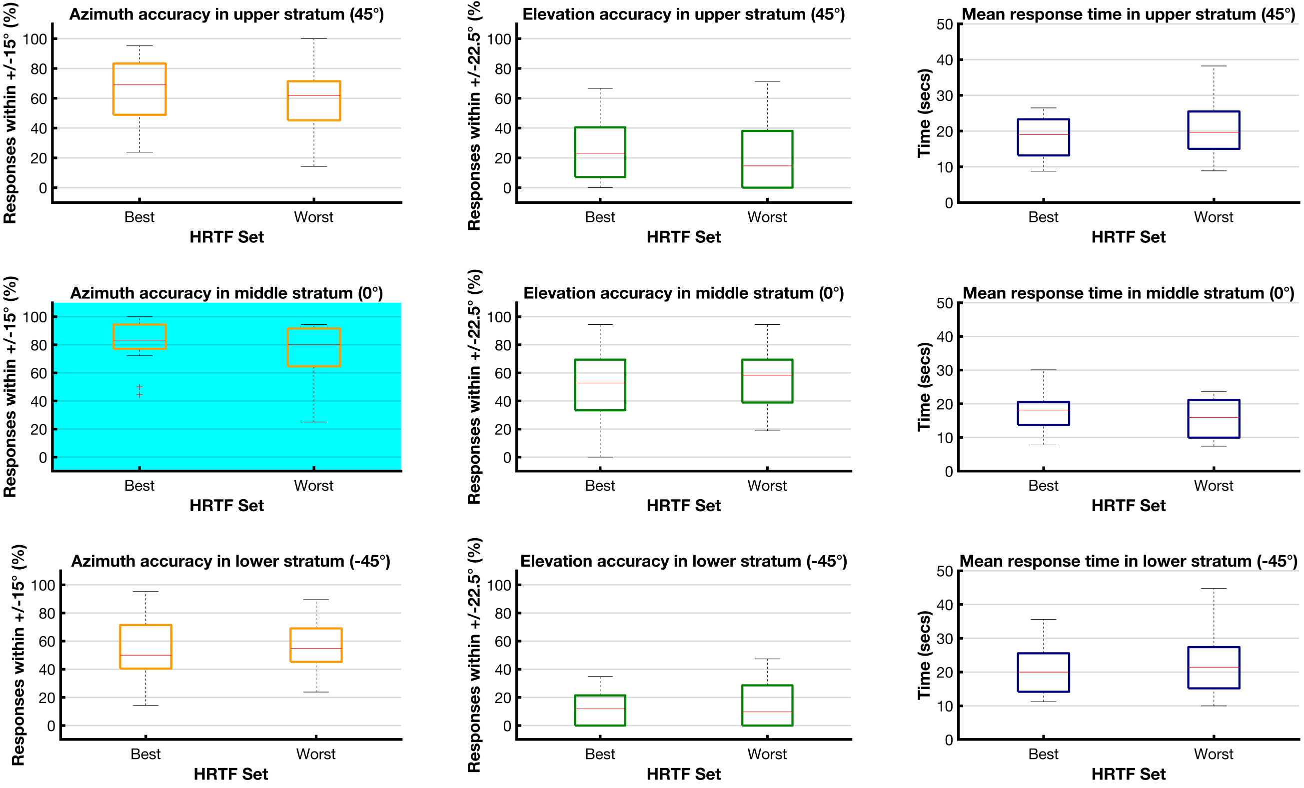


Figure 5. Distributions of participant azimuth/elevation success rates and mean response times, by HRTF preference*

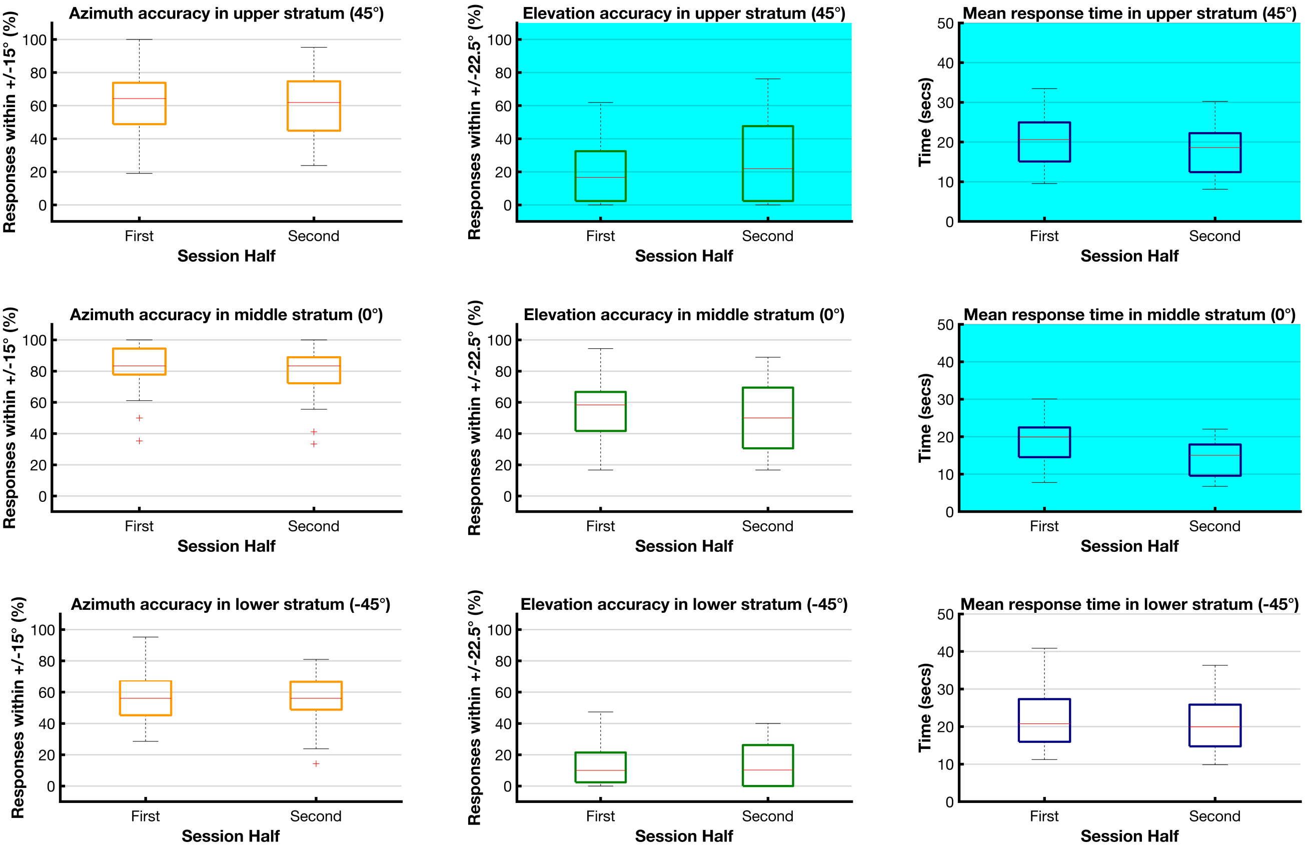


Figure 6. Distributions of participant azimuth/elevation success rates and mean response times, by sequence*

** Plots with blue background indicate significant difference between distributions* selection routine might therefore be validly applied in tandem with a structured pre-exposure phase to optimise perceptual experience.

It should also be noted that no meaningful statistical relationship was found between participants' HRTF ranking consistency and localisation performance. Significantly improved elevation accuracy was found in the 'Systematic Disagreement' group for sources on the horizontal plane. However, this apparent strength is actually a by-product of that group returning a greater overall proportion of responses that neglected vertical localisation and remained overly focussed at 0° elevation. The group was less likely to have noticed vertically displaced sources and performed particularly poorly in elevation accuracy at heights of 45° and -45°.

5.3 Usability

It is notable that the two most reliable raters judged themselves to be practised in binaural listening. However, there was no significant advantage to ranking consistency found through statistical analysis of musical training, headphone listening habits or prior binaural exposure. Moreover, some of those with only occasional and even no binaural experience were able to achieve fair or good levels of repeatability.

5.4 Efficiency

An average completion time between 12 and 13 minutes can be regarded as within the realms of an acceptable duration for single-time calibration of a 3D end-user system.

6. CONCLUSION

A new generation of headsets is enabling various forms of auditory mixed reality with interactive spatial sound. The efficacy of those binaural experiences will be at least partly impacted by the degree to which rendered scenes fit the perceptual profile of individual users. To date, there is no established method for personalising the HRTFs deployed for end-users of audio virtual or augmented reality in mobile contexts. The approach outlined here begins to address this issue and its first iteration has shown promising outcomes against previously identified evaluation criteria.

Acknowledgments

This work was supported by EPSRC and AHRC under the grant EP/L01632X/1 (Centre for Doctoral Training in Media and Arts Technology).

7. REFERENCES

[1] D. R. Begault, *3D sound for virtual reality and multimedia*, 1st ed. London, UK: Academic Press Limited, 1994.

[2] B. B. Bederson, "Audio augmented reality: a prototype automated tour guide," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Denver, Colorado, USA, 5 1995, pp. 210–211.

[3] M. Cohen, S. Aoki, and N. Koizumi, "Augmented audio reality : telepresence / VR hybrid acoustic environments," in *IEEE International Workshop on Robot and Human Communication*, Tokyo, Japan, 11 1993, pp. 361–364.

[4] N. Sawhney and C. Schmandt, "Nomadic Radio: speech and audio interaction for contextual messaging in nomadic environments," *ACM Transactions on Computer-Human Interaction*, vol. 7, no. 3, pp. 353–383, 2000.

[5] Y. Vazquez-Alvarez, M. P. Aylett, S. A. Brewster, R. Von Jungenfeld, and A. Virolainen, "Designing interactions with multilevel auditory displays in mobile audio-augmented reality," *ACM Transactions on Computer-Human Interaction*, vol. 23, no. 1, pp. 1–30, 2016.

[6] M. McGill, S. Brewster, D. McGookin, and G. Wilson, "Acoustic transparency and the changing soundscape of auditory mixed reality," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Honolulu, Hawaii,USA, 4 2020, pp. 1–16.

[7] A. N. Nagele, V. Bauer, P. G. T. Healey, J. D. Reiss, H. Cooke, T. Cowlishaw, C. Baume, and C. Pike, "Interactive audio augmented reality in participatory performance," *Frontiers in Virtual Reality*, vol. 1, no. February, pp. 1–14, 2021.

[8] R. Shukla, "Voice-led interactive exploration of audio," BBC Research and Development, Tech. Rep. November, 2019.

[9] M. Geronazzo, A. Rosenkvist, D. S. Eriksen, C. K. Markmann-Hansen, J. Køhlert, M. Valimaa, M. B. Vittrup, and S. Serafin, "Creating an audio story with interactive binaural rendering in virtual reality," *Wireless Communications and Mobile Computing*, vol. 2019, pp. 1–14, 2019.

[10] A. Roginska, "Binaural audio through headphones," in *Immersive sound: The art and science of binaural and multi-channel audio*, 1st ed., A. Roginska and P. Geluso, Eds. New York, NY, USA: Routledge, 2018, ch. 4, pp. 88–123.

[11] M. Geronazzo, S. Spagnol, and F. Avanzini, "Estimation and modeling of pinna-related transfer functions," in *Proceedings of the 13th International Conference on Digital Audio Effects*, Graz, Austria, 9 2010, pp. 1–8.

[12] A. Meshram, R. Mehra, H. Yang, E. Dunn, J.-M. Franm, and D. Manocha, "P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Munich, Germany, 9 2014, pp. 53–61.

[13] S. Spagnol, "HRTF selection by anthropometric regression for improving horizontal localization accuracy," *IEEE Signal Processing Letters*, vol. 27, pp. 590–594, 2020.

[14] R. Pelzer, M. Dinakaran, F. Brinkmann, S. Lepa, P. Grosche, and S. Weinzierl, “Head-related transfer function recommendation based on perceptual similarities and anthropometric features,” *The Journal of the Acoustical Society of America*, vol. 148, no. 6, pp. 3809–3817, 2020.

[15] B. U. Seeber and H. Fastl, “Subjective selection of non-individual head-related transfer functions,” in *Proceedings of the 2003 International Conference on Auditory Display*, Boston, MA, USA, 7 2003, pp. 259–262.

[16] A. Roginska, G. H. Wakefield, and T. S. Santoro, “User selected HRTFs: reduced complexity and improved perception,” in *Undersea Human System Integration Symposium*, Providence, RI, USA, 7 2010, pp. 1–14.

[17] R. Shukla, R. Stewart, A. Roginska, and M. Sandler, “User selection of optimal HRTF sets via holistic comparative evaluation,” in *AES International Conference on Audio for Virtual and Augmented Reality*, Redmond, WA, USA, 8 2018, pp. 1–10.

[18] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, “Localization using nonindividualized head-related transfer functions,” *The Journal of the Acoustical Society of America*, vol. 94, no. 1, pp. 111–123, 1993.

[19] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, “Binaural technique: Do we need individual recordings?” *Journal of the Audio Engineering Society*, vol. 44, no. 6, pp. 451–469, 1996.

[20] D. R. Begault, E. M. Wenzel, and M. R. Anderson, “Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source.” *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[21] A. Roginska, T. S. Santoro, and G. H. Wakefield, “Stimulus-dependent HRTF preference,” in *129th Audio Engineering Society Convention*, San Francisco, CA, USA, 11 2010, pp. 1–11.

[22] Y. Wan, A. Zare, and K. Mcmullen, “Evaluating the consistency of subjectively selected head-related transfer functions (HRTFs) over time,” in *AES 55th International Conference: Spatial Audio*, Helsinki, Finland, 8 2014, pp. 1–8.

[23] D. Schönstein and B. F. G. Katz, “Variability in perceptual evaluation of HRTFs,” *Journal of the Audio Engineering Society*, vol. 60, no. 10, pp. 783–793, 2012.

[24] A. Andreopoulou and A. Roginska, “Evaluating HRTF similarity through subjective assessments: factors that can affect judgment,” in *Joint 40th International Computer Music Conference & 11th Sound and Music Computing Conference*. Athènes, Greece: Michigan Publishing, 9 2014, pp. 1375–1381.

[25] A. Andreopoulou and B. F. G. Katz, “Investigation on subjective HRTF rating repeatability,” in *140th Audio Engineering Society Convention*, Paris, France, 6 2016, pp. 1–10.

[26] C. Kim, V. Lim, and L. Picinali, “Investigation into consistency of subjective and objective perceptual selection of non-individual head-related transfer functions,” *Journal of the Audio Engineering Society*, vol. 68, no. 11, pp. 819–831, 2020.

[27] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,” *Journal of the Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.

[28] R. Shukla, I. T. Radu, M. Sandler, and R. Stewart, “Real-time binaural rendering with virtual vector base amplitude panning,” in *AES International Conference on Immersive and Interactive Audio*, York, UK, 3 2019, pp. 1–10.

[29] V. Pulkki, “Evaluating spatial sound with binaural auditory model,” in *Proceedings of the 2001 International Computer Music Conference*. Havana, Cuba: Michigan Publishing, 9 2001, p. 73–76.

[30] R. Baumgartner and P. Majdak, “Modeling localization of amplitude-panned virtual sources in sagittal planes,” *AES: Journal of the Audio Engineering Society*, vol. 63, no. 7-8, pp. 562–569, 2015.

[31] O. Warusfel, “Listen HRTF database,” 2003. [Online]. Available: http://recherche.ircam.fr/equipes/salles/listen/index.html

[32] B. F. G. Katz and G. Parseihian, “Perceptually based head-related transfer function database optimization,” *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. EL99–EL105, 2012.

[33] V. Hansen and G. Munch, “Making recordings for simulation tests in the Archimedes project,” *Journal of the Audio Engineering Society*, vol. 39, no. 10, pp. 768–774, 1991.

[34] H. Bahu, T. Carpentier, M. Noisternig, and O. Warusfel, “Comparison of different egocentric pointing methods for 3D sound localization experiments,” *Acta Acustica united with Acustica*, vol. 102, no. 1, pp. 107–118, 2016.

[35] K. A. Hallgren, “Computing inter-rater reliability for observational data: an overview and tutorial,” *Tutorials in Quantitative Methods for Psychology*, vol. 8, no. 1, pp. 23–34, 2012.

[36] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*, revised ed. Cambridge, Massachusetts: MIT Press, 1997.