DOI: 10.61782/fa.2023.0744

AN EXPERIMENT REPLICATION ON DYNAMIC 3D SOUND LOCALISATION USING AUDITORY VIRTUAL REALITY

María Cuevas-Rodríguez[1]* **Daniel González-Toledo**[1]
Pablo Gutierrez-Parera[1,2] **Arcadio Reyes-Lecuona**[1]
[1]Telecommunication Research Institute (TELMA), Universidad de Málaga, Málaga, Spain
[2]Institute of Telecommunications and Multimedia Applications (ITEAM),
Universitat Politècnica de València, Spain

ABSTRACT

Dynamic cues due to self-motion are important in the perception of sound, as they allow individuals to increase their ability to localise sound sources. In a previous study, Gaveau et al. reported how spontaneous head movements improved 3D sound localisation by comparing static vs. active listening postures. They playback real sound sources with a portable loudspeaker placed at different positions relative to the listener's head. This contribution presents a partial replication of that experiment in a virtual environment using the 3D Tune-In Toolkit, an open-source C++ library for real-time binaural audio rendering. The anechoic acoustic path was rendered using a generic HRTF of a dummy head from the ARI database, and the reverberation of the room was reproduced with first-order Ambisonics-encoded BRIRs measured with an identical dummy head. The same room was also modelled in 3D for visual rendering. The results obtained in the virtual test carried out with headphones show findings comparable to those obtained by Gaveau et al. with real loudspeakers, enabling the validation of the virtual version of the experiment, although some differences appear regarding certain conditions. In addition, the importance of training in the virtual environment has been explored, indicating that its influence on performance is significant.

Keywords: *Binaural Rendering, Sound Localisation, Active Listening, Dynamic Listening, Virtual Reality*

1. INTRODUCTION

Perceptual sound localisation in humans can benefit from an active role of the listener or the sound sources. In such a case, the listener receives dynamic information that is not present in static listening where neither the sound source nor the listener is moving.

The binaural auditory system has traditionally been described as using three major cues for sound source localisation [1, 2]: the Interaural Time Differences (ITD), the Interaural Level Differences (ILD) and the spectral cues, contained in the Head Related Transfer Function (HRTF). Extensive work has been done on the study of auditory cues for localisation, but mainly in static conditions.

Dynamic variations of binaural perceptual cues such as ITD and ILD are beneficial for improving localisation accuracy in ambiguous spatial positions known as cones of confusion [3], and have been shown several times to clearly help in resolving front-back confusions [4, 5]. Along with dynamic interaural differences, dynamic variations of the spectrum can also help to improve the perception of elevated sources [6]. In addition to these acoustic dynamic cues, the own sensorimotion perception that produces self-motion cues also positively influences sound localisation [7]. There are additional benefits that improves the understanding of a sound scene, for example the use of auditory motion parallax to assess the relative distances of two sound sources [8]. Dynamic localisation is an active field of research [9, 10], also pursuing auditory modelling that can produce reliable predictions [11], and take into account proprioceptive information in the case of active listening [12].

Auditory virtual reality aims to faithfully recreate the

***Corresponding author*: mariacuevas@uma.es.*

acoustics of the real world, so that a realistic and plausible sense of sound immersion is achieved [13]. Binaural 3D audio is the most widespread technique for generating sound in virtual reality [14]. Static auralization using this method obtains very good perceptual results [15, 16], but some problems might appear when performing interactive auralizations in real-time: computational cost, acoustic and propagation rendering, smooth behaviour in dynamic situations, customization of HRTFs [17, 18].

The design and execution of dynamic localisation perceptual tests is also a complex task due to the multiple variables involved and their interrelationships, which elicit the perceptual process of localisation [6, 19, 20]. An example is the learning process that takes place during a perceptual test, which is also influenced by the existence or not of feedback on the subject's responses [21]. In the case of using a virtual environment, there are also the interactions that can occur due to the influence of visuals such as rooms and pointers [22], or the binaural sound rendering engine. This, which could be a difficulty when doing perceptual tests, can also be considered an advantage because the virtual tools can be used to isolate the different variables under study while maintaining the controllability and reproducibility of the experiment.

This paper presents a partial replication of a perceptual experiment on dynamic localisation conducted by Gaveau et al. [23] in which they examine whether active listening changed spatial hearing performance using a playback system with real loudspeakers. Their results show how allowing listener movement improved 3D sound localisation, ameliorating accuracy and variability of the responses. In contrast to the Gaveau et al. study, in our experiment the sound is synthesised using our own virtual audio rendering algorithms, which are employed in the 3D Tune-In Toolkit [24] and the Binaural Rendering Toolbox BRT [25], which synthesise real-time binaural sound that can be reproduced by headphones. The objective of this experiment is to perform a validation test of our binaural rendering tool, exploring its capabilities for realistic and accurate reproduction of virtual 3D sound.

2. METHODS

2.1 Virtual renderer

The virtual auditory scene has been rendered using the 3D Tune-In Toolkit [24], an open-source C++ library for binaural spatialisation in real-time. An HRTF-convolution based simulation is carried out to generate the direct acoustic path, also including the simulation of sources placed in the near field, using the model presented by Romblom in [26]. This model represents an extension of conventional Head-Related Transfer Function (HRTF) processing, using a difference filter for ILD, which predicts the spectral differences between a source in the near-field and another source located at the same azimuth and elevation angles, but at a distance where the HRTF was measured. The difference filter is based on the Spherical Head Model (SHM) presented by Duda et al. [27]. The reverberation of the room is computed separately using a virtual first order Ambisonic approximation [28]. The HRTF used in this study was that corresponding to the Neumann KU100 dummy head registered in the ARI database [29]. No headphones equalization has been included. Finally, the Binaural Room Impulse Responses (BRIRs) used to synthesise the reverberation were measured also employing a Neumann KU100 manikin and in the same room where the experiment took place.

2.2 Participants, apparatus and stimuli

Eight participants (all males) were recruited among students and researchers of the School of Telecommunication Engineering in the University of Malaga. All of them were right-handed and self-reported normal hearing. All procedures were reviewed and approved by the Ethical Committee for Research in Malaga (*Comité Ético de Experimentación de la Universidad de Málaga*).

Participants wore an Oculus Rift HMD for the visuals and a pair of Sennheiser HD600 headphones to reproduce the binaural audio. A Focusrite Scarlett 2i2 USB audio interface was used to play back the stimuli. An application was developed specifically for the experiment using Unity. This application renders the visual scene of the experiment, gets the tracker data from the Oculus Rift, and is used to configure all the visual and audio scenes. It communicates via OSC with the 3DTI Toolkit, which renders the audio scene.

A train of three 0.5s pulses of pink noise, separated by 0.5s silence, and spatialised using the 3D Tune-In AudioToolkit [24] was used as the stimulus.

2.3 Procedure

The experiment carried out takes into account an *active dynamic listening* situation, where the dynamic condition is given because listeners can move their head and obtain dynamic cues even from static sound sources [12]. In other words, the situation is dynamic because the listeners are active (active condition), but not the sources, which in this case do not move.

During the experiment, participants were seated in a chair, wearing the HMD, the headphones and holding one of the Oculus Rift controllers with their right hand. The Oculus Rift presented a visual simulation of the actual room they were sitting in. Participants' task was to listen to a sound and localise it, as accurately as possible, by indicating the position with the hand-held controller. Listeners performed the task under two listening conditions: static or active. During the active listening condition, participants had to move their heads while the sound is being played. For the static listening condition, they had to stay still until the sound has finished. In both conditions, participants had to indicate the position of the source once it has finished. Participants were told that the sound could be delivered at any position around them but with a maximum distance that allowed them to reach with their hand, without getting up from the chair. In fact, the virtual sound source was randomly positioned in 12 predetermined positions relative to the listener's head, resulting from the combination of 3 different distances (35 cm, 55 cm and 75 cm), and 4 different azimuths (30°, 150°, 210° and 330°) and 0° of elevation. This procedure and the source positions were a replication of the reference experiment developed by Gaveau et al. [23].

The experiment was structured in 2 sessions, each of them composed of 8 blocks with 12 trials. In each trial, the sound source was located in one of the 12 positions described previously. Each session had the two listening conditions, 4 blocks with static condition and 4 blocks with the active one. There was a break of 10 minutes between sessions. In addition, between the two sessions and after the break, there was a *training block*. This block has the same structure as the previous ones, but in this *training block*, participants had visual feedback, indicating, after their guess, which was the position of the sound source. Therefore, the sessions delimit another condition under study, pre-training for session 1 and post-training for session 2. The total number of stimuli was 96 for each session, with an average duration of 30 min for each session.

3. RESULTS

Figure 1 shows the bird's eye view and lateral view of all the participant's responses (guessed positions), for the static listening condition (blue dots) and the active listening condition (orange dots), averaged across trials for each participant. Black dots represent the target positions (the twelve virtual positions of the sources). In order to see the effect of the training, results have been separated into two different sessions. Pre-training session (session 1) includes trials performed before the training block and post-training session (session 2) the trials after the training. Thus, each dot represents the average of all the responses of one participant for one listening condition (static or active) and one target position, i.e. each point is the average of four responses of a participant.

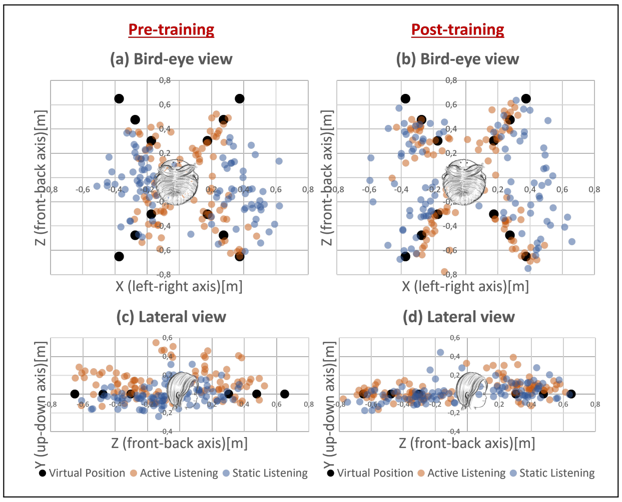

Figure 1. Bird's-eye and lateral view of all virtual sources positions (black dots) and responses averaged per participants, for static listening (blue dots) and active listening (orange dots). Responses have been separated by sessions: (a) and (c) Pre-training session, (b) and (c) Post-training session.

Figure 2 shows the responses for the different distances: near (0.35 m), middle (0.55 m) and far (0,75 m), also separated by sessions.

To study the effect of the listening condition in azimuth and elevation, the constant absolute error, which describes the accuracy of the performance, have been calculated using the Equation 1, where p indicates a specific participant. M is the number of participant responses for an specific experimental condition (target virtual position and listening condition); and N the number of different targets. For this study, $M = 4$ and $N = 12$. r is the the response of a participant for a target j, and k is the position (azimuth or elevation in degrees) of the target j.

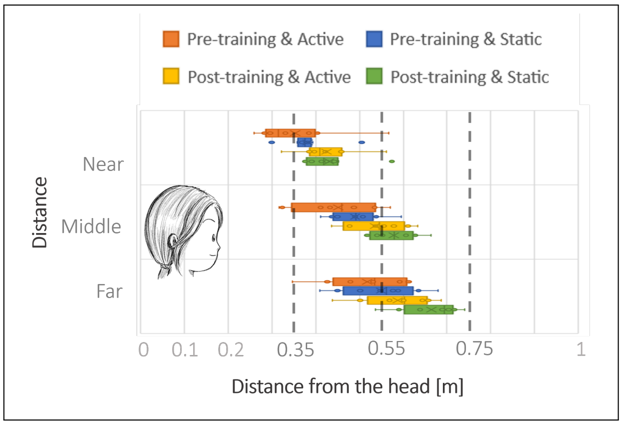

Figure 2. Distance responses for static and active listening condition, pre and post-training sessions.

$$C(p) = \frac{1}{M \cdot N} \sum_{j=1}^{N} \sum_{i=1}^{M} |(r_{i,j}(p) - k_j)| \tag{1}$$

Results have been analysed as a function of the listening condition (static and active), Z-axis position (listener's front and back) and training (pre-training and post-training); using a 3-ways ANOVA. The analysis of azimuth absolute constant error revealed significant differences for the session (training condition) $(F(1,7) = 19.451, p = 0.03)$ and listening condition $(F(1,7) = 24.2, p = 0.02)$ and no effect of the Z-axis position $(F(1,7) = 1.566, p = 0.251)$. The azimuth absolute constant error averaged for all participants is shown in Figure 3(a). The left side of the graph shows results for pre-training session and the right side for post-training session. Blue bars are the averaged errors for the static listing condition, orange bars represent the active listening condition. Results are grouped in the horizontal axis according to the target Z-axis position (front or back). These results show how the absolute constant errors in azimuth for the active listening condition were reduced compared to the static listening condition, both in the front and back positions and in both sessions. In addition, errors are reduced after training. The analysis of marginal means (corrected with Bonferroni) reveals an effect of the listening condition without training (pre-training session), with significant differences for the back $(p = 0.004)$ but not for the front $(p = 0.087)$. Similarly, after training (post-training session), significant differences can be found in the front $(p = 0.032)$ but not in the back section $(p = 0.142)$.

Absolute constant errors in elevations are shown in the same way in Figure 3(b). In the case of elevation, the error increases in the active listening conditions. ANOVA revealed that all main effects are significant: training $(F(1,7) = 14.572, p = 0.007)$, listening condition $(F(1,7) = 7.836, p = 0.027)$ and Z-axis position $(F(1,7) = 8.546, p = 0.022)$. However, the analysis of marginal means (with Bonferroni correction) reveals a significant difference of the listening conditions only in pre-training session and front position $(p = 0.015)$.

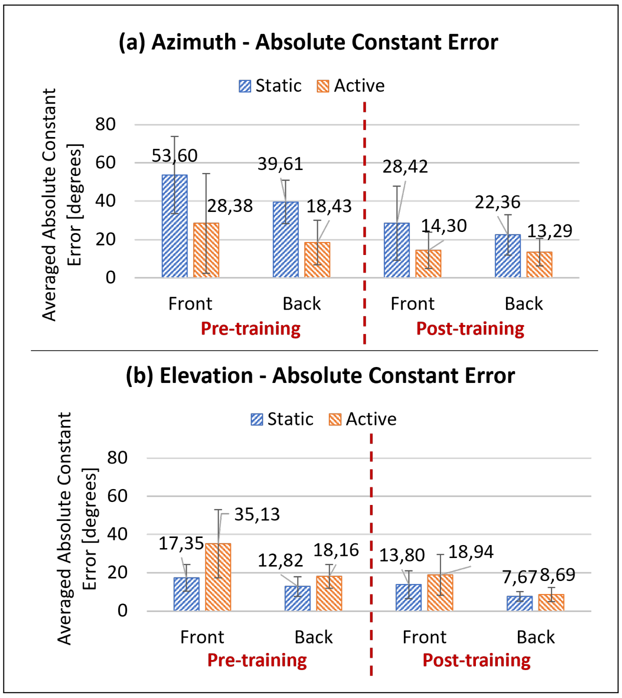

Figure 3. Absolute constant error for azimuth (a) and elevation (b), averaged for all participants. All errors bar are 95%CIs.

In addition, the 3D error has been calculated to quantify the overall sound localisation performance. Equation 2 shows how to calculate this error, using the same variables as in the Equation 1, but in this case $\overline{r_{i,j}(p)}$ and $\overline{k_j}$ are vectors that represents the participant response position and the target position in Cartesian coordinates.

$$D(p) = \sqrt{\frac{1}{M \cdot N} \sum_{j=1}^{N} \sum_{i=1}^{M} (\overline{r_{i,j}(p)} - \overline{k_j})^2} \tag{2}$$

Considering all the collected data, and in the same way as previously with the absolute constant error, we performed a 3-way ANOVA to study the effect of the different factors: training (pre-training session vs post-training session), listening condition (static and active) and Z-axis position (front and back). Results revealed a significant effect of training (session) $(F(1,7) = 110.716, p < 0.001)$ and listening condition $(F(1,7) = 52.775, p < 0.001)$, but no significant effect of the Z-axis position $(F(1,7) = 3.1, p = 0.122)$. The 3D averaged error for all participants are show in Figure 4, where it can be seen how the overall error were reduced significantly for the active as compared to the static listening condition, in both positions front and back. In addition, a reduction of the errors appears in the post-training session comparing with the pre-training one. The analysis of marginal means (Bonferroni corrected) reveals a significant effect of the listening condition, in the pre-training session for both front ($p = 0.002$) and back ($p = 0.002$) positions. After the training (post-training session), the effect of the listener condition is lower, with no significant differences in neither front ($p = 0.098$) nor back ($p = 0.574$) positions.

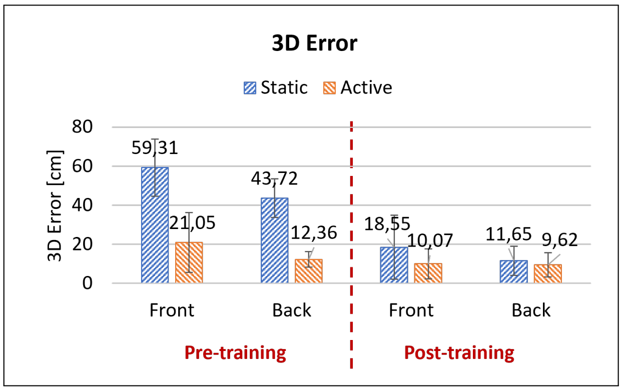

Figure 4. Overall error, averaged for all participants. All errors bar are 95%CIs. Data has been separated by sessions (pre-training and post-training), listening condition (static and active) and Z-axis position (front and back).

4. DISCUSSION

In order to check the validity of our binaural rendering tool to produce a realistic and acceptable virtual sound in a dynamic situation, we have made a critical comparison of our results with those obtained in the partially replicated reference study by Gaveau et al. [23], and we can see that the results obtained in our experiment follow the same pattern. The similarity with the results of the partially replicated study is even greater if we look at the responses given after the training (post-training session) (Figure 1).

For the azimuth localisation results, in both experiments the listening condition (static or active) has a significant effect and less errors are observed in the dynamic case compared to the static case. In addition, a significant effect of the training is observed, reducing the error and giving after training average values more similar to those obtained by Gaveau et al. than before training. Besides that, the error reduction for the active case is greater without training than after training, since training improves the results and therefore without training there is more room for improvement. Exploring the results we can see that the reduction in azimuth error achieved by active listening is different for the front and back positions. In the pre-training session, the error difference (improvement) is significant in the case of back positions, but not in the front. On the contrary, in the post-training session, significant differences appear in the front but not in the back, as in the experiment of Gaveau et al. In the post-training session, the visual feedback on the source positions and their relationship with the environment may have generated prior knowledge that could bring the results of our virtual experiment closer to those of Gaveau et al. The influence of prior knowledge may help in sound localisation tests by improving the awareness of the source and/or direction spectrum, to differentiate between spectral cues resulting from the source properties and from the filtering by the pinnae [12]. That is, sound localisation is an ill-posed perceptual problem given that the spectrum of the signal at the eardrum results from the combination of two unknowns: the spectrum of the sound source and the specific direction HRTF [30]. This is accentuated in the case of not using an individualised HRTF, as is our case and not that of Gaveau. Then, the training would be helping to reduce this ill-posed problem of the perceived spectrum and thus bringing the results of both localisation tests closer together.

It is also important to highlight that azimuth error for sources in the front was found to be larger than that for sources in the back. Although our sound localisation accuracy is better in the front than in the back, this result was expected, as we have not considered front-back confusions separated from other localisation errors and we have observed more front-back confusions for sources in the front. This is consistent with Gaveau's results who performed a similar analysis without separating front-back confusion effects.

Surprisingly, in the case of elevation in our experiment, worse results are observed in active listening than in static listening, although the differences are not significant except for the pre-training front positions. Visually, these differences can be seen in Figure 1 (c) and (d), where we can also see how the results improve after training. Previous studies show that elevation estimation based on dynamic ITD only, improves for elevations greater than 30° above or below the horizontal plane [5]. In our case the sources are initially located with no elevation with respect to the head orientation, elevation 0°, therefore the non-individual HRTF could be disturbing this elevation localisation. Furthermore, there seems to be a dependence of elevation perception on stimulus bandwidth [31] and even a direct relationship with dynamic spectral cues [6]. According to the latter study, with an imperfect HRTF not tailored for the particular listener, "moderate movements" of between 16-32° may be needed to significantly reduce elevation errors. In our study, we did not monitor how far the amplitude of the listener's movements reached, but it is possible that for some subjects and stimuli the movements were below this threshold. These arguments lead us to think that the dynamic condition does not necessarily make the elevation results better for our experimental conditions, in fact the differences in our case are not significant. However, it remains unexplained why we have found improvements for the static case. We think that this could be due to the fact that we didn't calibrate the position of the participant's head at the beginning of each trial. We just asked them to look at the front (where they had to press a button), and regarding this position, we placed the sound source relative to the listener (i.e., in the 0° elevation plane relative to the head position). This means that the horizontal plane where the sources were placed was tilted more than an absolute horizontal plane.

The distance perception results, depicted in Figure 2, are again similar to those of Gaveau et al. and are slightly more similar after training, with a noticiable externalisation of sources, according to the participants' responses. Neither of the two experiments seem to have a significant influence of distance on the results, and a well-known phenomenon of over-estimation of near distances and under-estimation of far distances [32] is also observed in both studies.

For the overall 3D error, the listening condition (static or active) which is only partially significant in the Gaveau et al. study is highly significant in our study, with a clear improvement in localisation due to active listening. It should be noted that the elevation errors commented before can have an influence on these overall 3D error results in our study, not present in the Gaveau et al. study.

In addition, it should be considered that there might be additional differences between the Gaveau et al. study and our experiment due to small procedural details such as the possible effect on listening of the Head Mounted Display in the Gaveau et al. experiment or the possible influence of the visual environment (Gaveau et al. used a black-no environment virtual space, ours used a virtual 3D laboratory room) [22, 33].

The training effect discussed above can be further influenced by an additional learning process about the spectral content of the sound source [34]. The interrelation of both processes is a complex phenomenon in which the user adapts to the playback system [35]. It would be interesting to minimise the effects of unfamiliarity of the source spectrum (e.g. by using natural sounds instead of synthetic noises as stimuli) as well as reduce the effects of user adaptation to the system (e.g. by using individual HRTFs with headphone response corrections). In this way we could better identify possible effects on localisation related to the chosen binaural rendering approaches.

5. CONCLUSIONS AND FUTURE WORK

Taking as reference a study on dynamic source localisation in 3D with sound reproduction through loudspeakers [23], a partial replication of this dynamic perceptual experiment has been carried out but with auditory virtual reality using our binaural rendering algorithms, using our binaural rendering algorithms [24, 25].

Our results show resemblance with those obtained by Gaveau et al., with slight differences that can be attributed to the procedure of the experiment or to the reproduction system. In both experiments, the influence of the active listening is significant and its positive influence is found for the different variables studied, except for the case of elevation, where there is a worsening. Furthermore, it has been observed that the effect of a training process brings the results of our test with virtual auditory closer to the results of the reference test with real loudspeakers.

The perceptual experiment presented here can also be considered as a pilot study to be repeated with more participants and adding more variables under study. As future work it would be interesting to include the influence of the visual environment, employ different sound stimuli including natural sounds to explore the realism of the virtual sound simulation, include individual HRTFs (and specifically check elevation performance), as well as monitor and register listener movements. It should be noted that the described training effect is additionally influenced by the learning effect of the first block of trials. It would be desirable to establish control groups in future tests to separate the two effects.

The findings resulting from this comparison study will allow us to conduct subsequent perceptual experiments with a quantified degree of confidence, while the exploration of the differences will allow us to refine our binaural synthesis tool and algorithms, in a quest to create a full virtual audio laboratory.

6. ACKNOWLEDGMENTS

This study has been supported by SONICOM `www.sonicom.eu`, a project funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 101017743, the Spanish National Project SAVLab, under grant No. PID2019-107854GB-I00, funded by MCIN/AEI/10.13039/501100011033/FEDER UE, and with the postdoctoral program APOSTD (grant number CIAPOS/2021/132) by the Generalitat Valenciana and European Social Fund.

7. REFERENCES

[1] J. C. Middlebrooks and D. M. Green, "Sound Localization by Human Listeners," *Annual Review of Psychology*, vol. 42, pp. 135–159, jan 1991.

[2] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *The Journal of the Acoustical Society of America*, vol. 111, p. 2219, may 2002.

[3] H. Wallach, "The role of head movements and vestibular and visual cues in sound localization.," *Journal of Experimental Psychology*, vol. 27, pp. 339–368, oct 1940.

[4] W. R. Thurlow, J. W. Mangels, and P. S. Runge, "Head Movements During Sound Localization," *The Journal of the Acoustical Society of America*, vol. 42, pp. 489–493, aug 1967.

[5] S. Perrett and W. Noble, "The contribution of head motion cues to localization of low-pass noise," *Perception and Psychophysics*, vol. 59, pp. 1018–1026, jan 1997.

[6] K. I. McAnally and R. L. Martin, "Sound localization with head movement: Implications for 3-d audio displays," *Frontiers in Neuroscience*, vol. 8, p. 210, aug 2014.

[7] C. Kim, R. Mason, and T. Brooke, "Head movements made by listeners in experimental and real-life listening activities," *AES: Journal of the Audio Engineering Society*, vol. 61, pp. 425–438, jul 2013.

[8] B. G. Shinn-Cunningham, S. Santarelli, and N. Kopco, "Tori of confusion: Binaural localization cues for sources within reach of a listener," *The Journal of the Acoustical Society of America*, vol. 107, pp. 1627–1636, mar 2000.

[9] H. Pöntynen and N. H. Salminen, "Resolving front-back ambiguity with head rotation: The role of level dynamics," *Hearing Research*, vol. 377, pp. 196–207, jun 2019.

[10] G. McLachlan, P. Majdak, J. Reijniers, M. Mihocic, and H. Peremans, "Dynamic spectral cues do not affect human sound localization during small head movements," *Frontiers in Neuroscience*, vol. 17, p. 75, feb 2023.

[11] N. Ma, T. May, and G. J. Brown, "Exploiting Deep Neural Networks and Head Movements for Robust Binaural Localization of Multiple Sources in Reverberant Environments," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 12, pp. 2444–2453, 2017.

[12] G. McLachlan, P. Majdak, J. Reijniers, and H. Peremans, "Towards modelling active sound localisation based on Bayesian inference in a static environment," *Acta Acustica*, vol. 5, p. 45, oct 2021.

[13] M. Geronazzo and S. Serafin, *Sonic Interactions in Virtual Environments*. Human–Computer Interaction Series, Cham, Switzerland: Springer International Publishing, 2023.

[14] J. Blauert, *The technology of binaural listening*. Springer, 2013.

[15] M. Blau, A. Budnik, M. Fallahi, H. Steffens, S. D. Ewert, and S. van de Par, "Toward realistic binaural auralizations – perceptual comparison between measurement and simulation-based auralizations and the

real room for a classroom scenario," *Acta Acustica*, vol. 5, p. 8, jan 2021.

[16] H. Mi, G. Kearney, and H. Daffern, "Perceptual Similarities between Artificial Reverberation Algorithms and Real Reverberation," *Applied Sciences*, vol. 13, p. 840, jan 2023.

[17] M. Vorländer, *Auralization*. RWTHedition, Cham, Switzerland: Springer International Publishing, second ed., 2020.

[18] N. Raghuvanshi and H. Gamper, "Interactive and Immersive Auralization," in *Sonic Interactions in Virtual Environments* (M. Geronazzo and S. Serafin, eds.), pp. 77–113, Cham, Switzerland: Springer International Publishing, 2023.

[19] S. Carlile and J. Leung, "The Perception of Auditory Motion," *Trends in Hearing*, vol. 20, jan 2016.

[20] D. Morikawa, Y. Toyoda, and T. Hirahara, "Head movement during horizontal and median sound localization experiments in which head-rotation is allowed," in *Proceedings of Meetings on Acoustics*, vol. 19, pp. 050141–050141, 2013.

[21] P. Majdak, M. J. Goupell, and B. Laback, "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, Perception, and Psychophysics Psychophysics*, vol. 72, pp. 454–469, feb 2010.

[22] A. Ahrens, K. D. Lund, M. Marschall, and T. Dau, "Sound source localization with varying amount of visual information in virtual reality," *PLOS ONE*, vol. 14, p. e0214603, mar 2019.

[23] V. Gaveau, A. Coudert, R. Salemme, E. Koun, C. Desoche, E. Truy, A. Farnè, and F. Pavani, "Benefits of active listening during 3D sound localization," *Experimental Brain Research*, vol. 240, pp. 2817–2833, nov 2022.

[24] M. Cuevas-Rodriguez, L. Picinali, D. Gonzalez-Toledo, C. Garre, E. de la Rubia-Cuestas, L. Molina-Tanco, and A. Reyes-Lecuona, "3D tune-in toolkit: An open-source library for real-time binaural spatialisation," *PLoS ONE*, vol. 14, p. e0211899, mar 2019.

[25] D. Gonzalez-Toledo, L. Molina-Tanco, M. Cuevas-Rodriguez, P. Majdak, and A. Reyes-Lecuona, "The Binaural Rendering Toolbox. A virtual laboratory for reproducible research in psychoacoustics," in *Forum Acusticum 2023*, (Turin, Italy), 2023.

[26] D. Romblom and B. Cook, "Near-Field Compensation for HRTF Processing," in *125th Convention Audio Engineering Society*, (San Francisco, USA), Audio Engineering Society, oct 2008.

[27] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Am.*, vol. 104, no. 5, pp. 3048–3058, 1998.

[28] I. Engel, H. Craig, S. V. Amengual Garí, P. W. Robinson, and L. Picinali, "Perceptual implications of different Ambisonics-based methods for binaural reverberation," *J. Acoust. Soc. Am.*, vol. 149, no. 2, pp. 895–819, 2021.

[29] Acoustics Research Institute of the Austrian Academy of Sciences, "ARI HRTF-Database," 2013. Available at `https://www.oeaw.ac.at/en/isf/das-institut/software/hrtf-database`.

[30] R. Ege, A. J. V. Opstal, and M. M. Van Wanrooij, "Accuracy-Precision Trade-off in Human Sound Localisation," *Scientific Reports*, vol. 8, p. 16399, nov 2018.

[31] T. Ashby, T. Brookes, and R. Mason, "Towards a head-movement-aware spatial localisation model: Elevation," in *21st International Congress on Sound and Vibration 2014, ICSV 2014*, vol. 4, pp. 2808–2815, 2014.

[32] A. J. Kolarik, B. C. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, "Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss," *Attention, Perception, and Psychophysics*, vol. 78, pp. 373–395, feb 2016.

[33] M. Cuevas-Rodriguez, D. L. Alon, S. W. Clapp, P. W. Robinson, and R. Mehra, "Evaluation of the effect of head-mounted display on individualized head-related transfer functions," in *International Congress on Acoustics*, vol. 2019-Septe, (Aachen, Germany), pp. 2635–2642, 2019.

[34] C. Mendonça, G. Campos, P. Dias, and J. A. Santos, "Learning Auditory Space: Generalization and Long-Term Effects," *PLoS ONE*, vol. 8, no. 10, 2013.

[35] L. Picinali and B. F. G. Katz, "System-to-User and User-to-System Adaptations in Binaural Audio," in *Sonic Interactions in Virtual Environments* (M. Geronazzo and S. Serafin, eds.), pp. 115–143, Cham, Switzerland: Springer International Publishing, 2023.