forum acusticum 2026
8-12 September · Graz, Austria

BASELINE LOCALISATION BIASES IN AUDIO-ONLY VIRTUAL REALITY PRIOR TO INVESTIGATING FLOOR REFLECTIONS

Max Væhrens[1], Stefania Serafin[2], Flemming Christensen[1], 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*

Abstract

In real environments, direct sound from a source is typically followed by a strong floor reflection whose level and time delay depend on the location of the source and the listener, and may potentially provide additional localisation cues. However, prior findings on reflection effects are mixed, and evidence for virtual reality (VR) applications and environments remain limited. This question is especially relevant for blind and visually impaired (BVI) users, who may rely heavily on accurate auditory spatial cues in audio-only VR for tasks such as orientation and locomotion involving footsteps, cane-impact and echolocation. We developed a dynamic VR sound localisation platform with spatialised rendering of direct sound and options for reflections. This paper reports baseline results from initial experiments using direct sound only. Across stimuli, responses showed significant compression toward the horizontal plane (systematic elevation bias) for controller responses and head pointing. Elevation was a strong predictor of signed localisation error, and sound stimulus significantly modulated the elevation bias slope. In contrast, no robust effect of HRTF was observed on the elevation-bias function, and no sound and HRTF interaction was found. These baseline results provide a reference for subsequent experiments testing whether adding floor reflections reduces elevation-bias compression in VR. We hypothesise that, relative to this baseline, adding floor reflections will reduce elevation-bias compression.

Keywords: *spatial audio, virtual reality, localisation, room acoustics*

Corresponding author: *maxv@es.aau.dk.*

1. Introduction

Virtual Reality (VR) experiences and games are able to provide a close to fully immersive experience by capturing the attention of the primary senses, namely vision and hearing. However, VR games remain largely inaccessible to blind and visually impaired (BVI) users because visual rendering dominates VR scenes, interaction and controls. The principal focus of the present efforts is to further develop options for audio-only VR games which offer convincing experiences to BVI gamers. For the present study, the purpose is to explore the importance of sound reflections, which may aid in substituting visual rendering. For instance, reverberation has been shown to improve externalisation rates of virtual speech stimuli [1], and noise, music, and environmental sounds [2].

The motivation for this work stems from the hypothesis that BVI users may require accurate auralisation. E.g. accurate binaural rendering of footsteps, echolocation, and/or white cane sounds may be crucial in perceiving the height of the virtual floor precisely during six-degree of freedom (6DoF) movement and while maintaining balance. It may be critical that the perception of the virtual floor is consistent with the physical floor, so that the tactile feedback while standing and moving corresponds with the auditory feedback. Furthermore, improving the BVI users' ability to localise important game objects in 3D space and perceive the 3D environment may be crucial to increase the feeling of immersion and responsiveness of audio-only games.

Prior studies have investigated the influence of reflecting surfaces on sound localisation. Rakerd and Hartmann found that reflections from a lateral surface decreased azimuthal localisation accuracy of impulsive tones compared to free field conditions, while reflections in the vertical axis had a smaller effect [3]. Between anechoic and a moderately reflective environment, Zahorik et al. found no significant differences between reversal rates and distance judgement was unaffected. This study concluded that when using individualised head-related transfer functions (HRTFs) in virtual auditory displays, accurate room reflections may perhaps be unnecessary for high localisation performance [4]. However, Guski found an increase in correct localisation of sound sources with added reflections from a floor surface, while reflections from a ceiling degraded localisation. Furthermore, the floor reflections especially improved sound source elevation detection [5]. In virtual spatialisation experiments, Reed and Maher showed that certain reflections from side walls aid in solving front-back localisation when using generalised HRTFs [6]. Kamizono et al. showed that reflected sound from side walls affect the front-back perception for high frequency narrow-band noise [7]. Wendt et al. investigated the influence of a simulated floor reflection on source height perception and found that speech signals were affected, whereas broadband noise was unaffected [8]. Begault et al. found that both early reflections and full reverberation compared to anechoic rendering improved azimuth judgement, while degrading elevation judgement for brief speech segments [1], however the experiment only included sound sources in the horizontal plane. Zamir et al. showed that localisation of speakers in a sound-treated room were significantly better (average improvements of 2° and 4° for azimuth and elevation, respectively) than virtual localisation using individual HRTFs, due to the virtual rendering being anechoic.

The conflicting reports in the literature pose a challenge in inferring which degree to use early reflections and reverberation when implementing virtual auralisation. In our preliminary observations, we observed that localisation of direct sound from sources located near the floor were generally perceived as being located vertically further above the correct position. These observations stemmed from an active dynamic sound localisation experiment using an audio-only VR test platform [9]. Perhaps, the upwards tendency may have been caused by factors such as using unknown generic HRTFs, open speakers built into the VR head-mounted display (HMD), or due to the frequency and tonal components of the stimuli. Yet, we hypothesise that the lack of a strong floor reflection may also contribute to the upwards tendency. This paper presents baseline results for localisation of direct sound. The experiment data are from a second iteration of the active dynamic sound localisation experiment [10].

2. Methods

The VR test platform facilitates an active dynamic sound localisation experiment based on a classical 'shooting down sound sources' task [11]. In the experiment the sound source was invisible except during training sessions, which served as familiarisation with the task and the unique sound. The VR controllers are rendered with a visible laser pointer to aid the participants in pointing the controller, thereby reducing requirement for good hand-eye motor control and variability [12], [13]. The participants could point and shoot with either of the VR controllers, whichever they preferred, by pulling the main trigger. The participants were instructed to find and shoot down sound sources one at a time at different locations in the space surrounding the participant. The sound stimuli were emitted continuously during each trial enabling utilisation of active dynamic listening to localise each source location. The virtual environment is illustrated in Figure 1 with a sample of fifteen sound sources visible for exemplary demonstration, as only a single sound source is active at any given time.

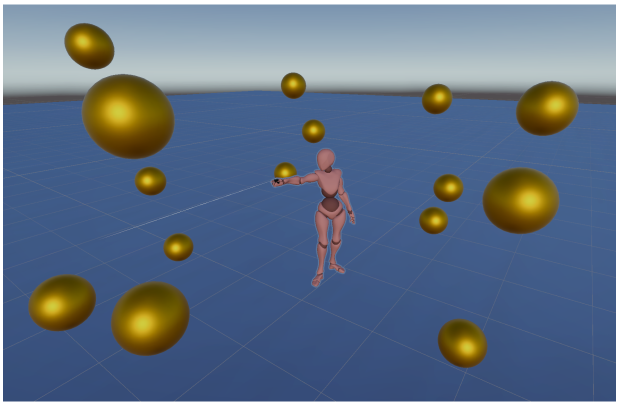

Figure 1: Screenshot of the third-person view of the virtual environment showing a virtual character in the centre of fifteen visible sound sources. The character holds and points a controller with a white laser pointer (ray) in the right hand.

The VR test platform was implemented in Unity and operates on a computer connected to a Meta Quest Pro headset. The platform enables testing of different source locations distributed on one or more spheres surrounding the subject; different HRTFs; various sound stimuli and duration thereof; pointing and aiming instructions; and audio-only and/or visual search and localisation tasks.

2.1. Protocol and stimuli

Four different sound stimuli were tested in eight sessions: white noise; recording of a female reading a fairytale in Danish; recording of a burning fire; and two sonar beeps, see [10] for a full description. The stimuli had duration between three seconds and twenty minutes (speech) and was looped when a trial lasted longer than the stimulus duration. White noise and speech were tested across four sessions on each participant's first appointment, and the latter two stimuli on a second appointment. A break was administered between each session and lasted approximately as long as each session (five to seven minutes) to reduce fatigue. Hence, each appointment including introduction and debrief lasted approx. an hour. Each stimulus was tested twice and the stimuli order for each appointment was balanced across subjects. Each session had a first block with visible sources for familiarisation, followed by three blocks with invisible sources and three unique HRTFs, and a final invisible block repeating the HRTF of the second block. Each block regardless of type comprised fifteen trials, each trial had a unique sound source position picked randomly. Presentation order of the HRTF blocks were also balanced for each session. Sound source positions were randomly generated for each block with an approximately uniform spherical distribution within a $\pm 45°$ elevation band, implemented to reduce learning of specific source positions.

2.2. Binaural rendering

Sound was spatialised using Steam Audio SDK for Unity using three non-individual head-related impulse response (HRIR) datasets measured on human subjects [14]. A fourth unique HRIR dataset was used for the familiarisation blocks to avoid training effects. The HRIR datasets were measured from 193 source positions and real-time linear interpolation was applied to obtain positions falling between measured positions. The Quest Pro HMD was connected wirelessly to a computer running the test platform through a standard WiFi-router. The spatialised sound was emitted using Beyerdynamic DT990 open-back headphones connected directly to the HMD. The participants were standing in the centre of an approximately 3 x 3 meter open area in a large sound-treated room and could move around freely with six degrees of freedom tracking. A guardian zone corresponding to the open area was set up in the HMD firmware for safety.

2.3. Subjects

The participants were fifteen students (four women, eleven men) whom all completed the first appointment, thereof five participants (all men) completed the second appointment on a separate day. The participants were all volunteers recruited through convenience sampling. All participants had self-reported normal vision and hearing status, and had various prior experience with VR, however no prior experience with localisation experiments.

2.4. Data collection and analysis

Signed elevation errors were computed as the signed angular difference between the sound source direction and the pointing direction of the controller as the response is given. These are used to indicate potential overshooting or undershooting behaviour ("localisation bias") relative to the source positions [13], [15]. Positive elevation indicates upward offset to the source and negative elevation indicates the contrary. The analysis was also run for head pointing directions (the forward facing direction of the VR headset) at the time of response as an exploratory step. Responses were classified as being within the cone-of-confusion if the absolute azimuth is less than 5° and out-of-cone above this threshold. A significance level of $\alpha = 0.05$ was used in the statistical analysis.

3. Results

A total of $n = 4800$ invisible trials were completed by the $N = 15$ participants across the 20 appointments. Single trials with controller azimuth above 150° were deemed as outliers and were removed from the dataset, as these trials were either an early discharge or a front-back reversal. As such, 6 trials (0.1% of all trials) were excluded from the analysis. Thereof, approx. 79% of controller responses and 71% of head pointing directions were placed within the cone-of-confusion of the sound source. Results for the head pointing measure are included in the analysis as initial analysis showed that participants naturally orient their head towards the sound sources, despite not being instructed to point their head towards the perceived source position when giving a response. These results indicate that the participants were able to use interaural cues to resolve the azimuth position of the sound sources in most cases.

3.1. Signed localisation bias across source elevations

Figure 2 shows signed elevation error as a function of grouped sound source elevation (from $-45°$ to $45°$ in steps of 10° groups) for controller responses and head pointing. Across all sound stimuli, signed elevation error depended strongly on source elevation. Sources presented below the horizontal plane were systematically localised too high, whereas sources presented above the horizontal plane were localised too low. This effect increased with distance from the horizontal plane, indicating a compression of the perceived elevation range towards the horizontal plane. One-sample t-tests revealed that the effect was particularly consistent for the white noise and speech stimuli. For white noise, significant positive elevation biases were observed for the four lowest elevation groups for both controller and head pointing (all $p < 0.004$, $d = 0.91$ to -1.62), while significant negative biases were observed at the highest elevation group (35° to 45°, controller: $p = 0.023$, $d = -0.66$; head: $p = 0.001$, $d = -1.10$). Similarly, speech produced significant positive biases for elevation groups below 0° and significant negative biases for groups above 15° (all $p < 0.005$, $d = 0.86$ to -2.71), with biases reaching approximately 25° at the lowest elevations and $-16°$ at the highest elevations.

The compression effect was most pronounced for the sonar-beep stimulus. Large positive biases were observed for the four lowest elevation groups and large negative biases for the three highest groups for both response methods (all $p < 0.023$, $d = 1.61$ to -6.42), reaching approximately 41° at low elevations and $-26°$ at high elevations.

For the fire stimulus, the same overall trend was present but effects were smaller and more variable. Significant deviations from zero were limited to a subset of elevation groups (5° to 45°), primarily for head-pointing responses (all $p < 0.043$, $d = 1.32$ to -2.59).

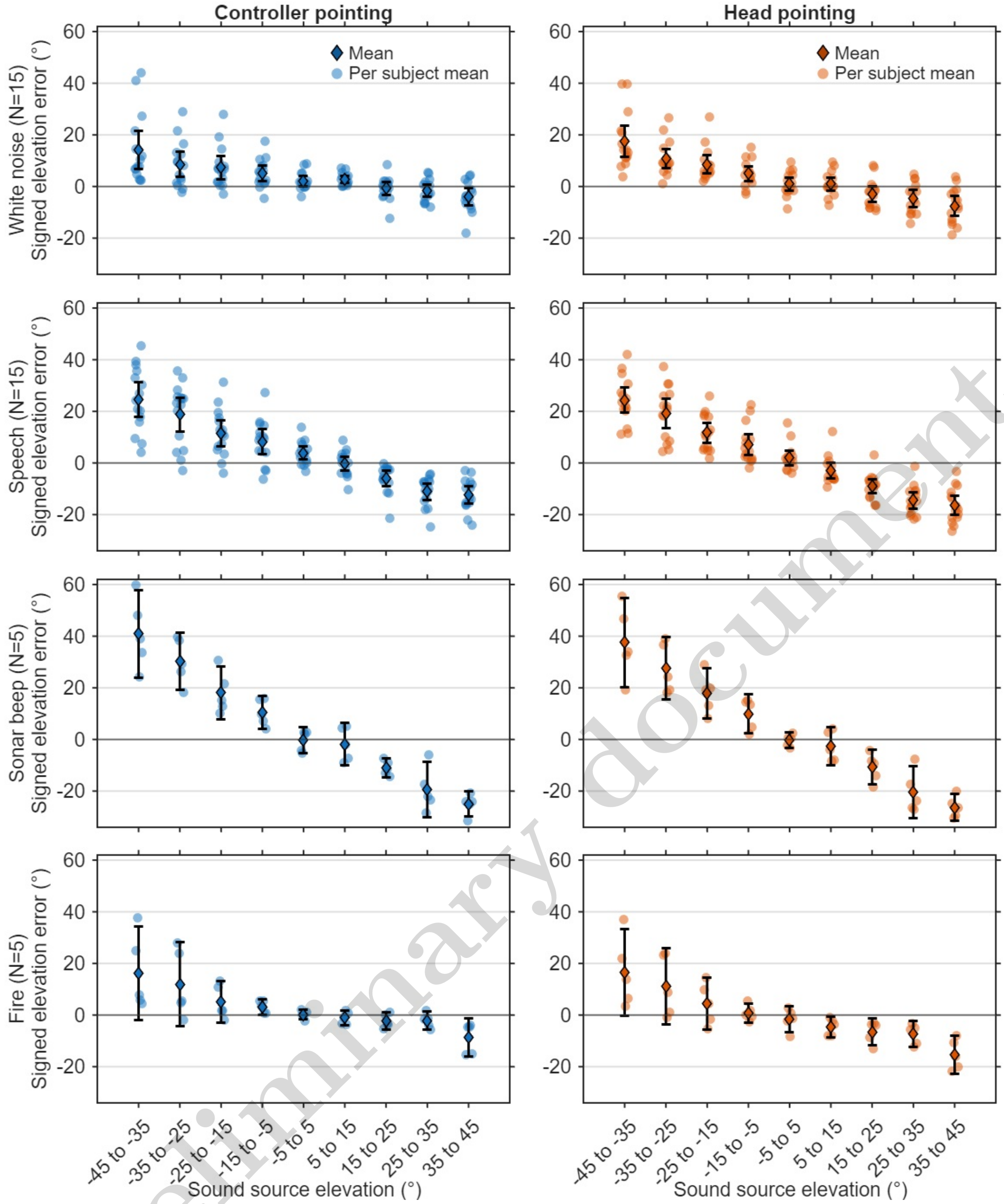

Figure 2: Signed elevation error plots for controller pointing (first column) and head pointing (second column) for a specific sound stimuli (each rows) and grouped by sound source elevation (x-axis). Diamonds indicate mean values across subjects and circles represents per subject means. Whiskers indicate 95% confidence interval.

3.2. Signed localisation bias across HRTF, sound, and elevation

To test whether HRTF rendering and source elevation (and their interaction) influenced elevation pointing bias, while also modelling sound stimulus effects, linear mixed-effects models predicting signed elevation error were fitted for controller and head pointing. Source elevation was mean-centred and included as linear and quadratic terms to capture, respectively, the overall rate and the curvature of the elevation-dependent pointing bias. The linear term captures the overall (average) rate of change in signed elevation bias per degree of source elevation. I.e., whether responses systematically undershoot or overshoot as the source moves away from the horizontal plane, and by how much per degree. The quadratic term captures curvature in the bias function. I.e., whether the undershoot/overshoot becomes much larger (or smaller) at extreme elevations than a straight-line relationship would predict, independent of the direction of the linear trend. Because each participant was tested with all three HRTFs within each sound stimulus session, sound stimuli and HRTFs were fully crossed within each subject. A random intercept for subject accounted for repeated measures. All four sound stimuli were included (N = 15 participants for white noise and speech, N = 5 for sonar beeps and fire sound, hence 4794 trials total). Fixed-effect significance was assessed via Type III F-tests and Bayes factors (BF_{10}, BIC-approximation).

HRTF showed no significant main effect on signed elevation error (controller: $F(2, 4776) = 2.19$, $p = 0.112$; head pointing $F(2, 4776) = 0.96$, $p = 0.381$), and no interaction with linear elevation (controller: $F(2, 4776) = 0.18$, $p = 0.839$; head: $F(2, 4776) = 0.05$, $p = 0.950$). The HRTF and quadratic elevation interaction reached nominal significance for controller pointing ($F(2, 4776) = 4.48$, $p = 0.011$), but not for head pointing ($F(2, 4776) = 2.44$, $p = 0.087$). However, the Bayesian model comparison favoured excluding HRTF-related terms ($BF_{10} < 0.01$), so there was stronger evidence for no meaningful HRTF effect overall. Hence, these results do not indicate that HRTF rendering shifted or changed the shape of the elevation-dependent bias.

Source elevation had a strong negative linear effect on signed pointing error for both controller ($\beta = -0.199$, $SE = 0.019$ (standard error), $t(4776) = -10.54$, $p < 0.001$) and head pointing ($\beta = -0.284$, $SE = 0.016$, $t(4776) = -18.22$, $p < 0.001$), reflecting a systematic undershoot at more eccentric elevations. The quadratic term was small and inconsistent across pointing measures (controller: $p = 0.074$; head: $p = 0.010$), and the corresponding Bayes factors indicated the full quadratic term set was not well supported ($BF_{10} \leq 0.001$), suggesting the bias function is close to linear over the tested elevation range. The linear elevation slope differed significantly by sound (controller: $F(3, 4776) = 181.8$, $p < 0.001$; head: $F(3, 4776) = 185.8$, $p < 0.001$). The slope was steepest for sonar beeps (controller: $\beta = -0.815$; head: $\beta = -0.786$) relative to white noise (controller: $\beta = -0.199$; head: $\beta = -0.284$), followed by speech (controller: $\beta = -0.472$; head: $\beta = -0.532$), while fire sound showed a comparatively shallow slope (controller: $\beta = -0.276$; head: $\beta - 0.350$) but a large overall downward shift in the intercept (controller: $\beta = -4.50$, $p < 0.001$; head: $\beta = -4.26$, $p < 0.001$), relative to white noise. The sound and HRTF interaction was not significant (controller: $F(6, 4782) = 0.10$, $p = 0.997$; head: $F(6, 4782) = 0.17$, $p = 0.985$), meaning the sound-related bias differences did not depend on which HRTF was used. Sound had a strong main effect (controller: ($F(3, 4776) = 8.67$, $p < 0.001$); head: ($F(3, 4776) = 8.37$, $p < 0.001$). As sonar beeps and fire were tested in a smaller subsample, sound-specific comparisons should be treated as exploratory.

4. Discussion

In this study, participants localised continuous virtual sound sources under active listening, allowing dynamic binaural cue sampling during head movements. Although this paradigm differs from prior static localisation paradigms using noise bursts and individual HRTFs, such as Tabry et al. [13], we observed a similar approximately linear relationship between signed elevation error and source elevation.

Despite the availability of dynamic cues, signed elevation error depended strongly on source elevation for both controller and head pointing. This elevation dependence was not meaningfully modulated by HRTF: we found no robust HRTF main effect, no reliable HRTF interaction with linear elevation, and no sound and HRTF interaction; Bayesian model comparisons also favoured excluding HRTF-related terms. These findings suggest that the observed elevation bias is driven by factors shared across conditions rather than by the specific HRTF set used here. Candidate contributors include response strategy; pointing interface characteristics; headphone rendering; and shared properties of the virtual acoustic environment. It also remains possible that the tested HRTFs were spectrally similar enough in the elevation-relevant cue regions that no differential bias emerged; broader and individualised HRTF sets should be tested in future work.

In contrast, sound stimulus strongly interacted with source elevation, indicating sound-specific differences in elevation bias (e.g., a steeper bias slope for speech relative to white noise). This is consistent with an interpretation that spectral and/or temporal properties of the stimuli [10] influence elevation judgements. In particular, the narrowband and tonal sonar beeps showed the steepest compression pattern. However, sonar and fire results should be interpreted cautiously because they were obtained for a smaller subsample of only five participants.

5. Conclusion

A virtual audio-only localisation experiment was trialled by fifteen seeing participants. The localisation task involved active dynamic listening in anechoic conditions. The experiment incorporated four unique continuously rendered sound stimuli, sound sources distributed around the listener and three non-individual head-related transfer functions. Across both controller responses and head-pointing, signed elevation error showed systematic compression toward the horizontal plane. Sound stimulus significantly modulated the elevation-bias slope, with stronger compression for stimuli with reduced broadband spectral content (most notably sonar beeps). By contrast, elevation bias did not show robust dependence on HRTF, and no interaction between sound and HRTF was observed. Thus, in this paradigm, elevation bias appears to be primarily associated with elevation and stimulus characteristics rather than by the specific non-individual HRTF used. Sound-specific comparisons involving sonar beeps and fire are exploratory due to a smaller sample size. Replication with larger balanced samples and individualised HRTFs is needed to determine whether subtle HRTF-dependent effects emerge under otherwise identical task conditions. The present findings define a no-reflection reference condition against which the localisation benefit of floor reflections can be tested.

6. Future Work

As the auralisation in the present study was anechoic, future experiments should include floor reflections and room acoustics for comparison with the present results. In the latter case, the importance of early reflections versus late reflections may be emphasised. Furthermore, the effect of accommodation to or familiarisation with specific virtual acoustic environments may also be studied [16]. Future experiments should also test a cohort of blind and visually impaired subjects, as this population may be more acute to differences in reflections or to the absence of reflections. These experiments should also consider the pointing and response method, as blindfolding has been shown to result in significantly lower absolute localisation errors [13].

7. Acknowledgement

We want to thank our collaborators Antonia Lina 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] 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," *J. Audio Eng. Soc.*, vol. 49, no. 10, pp. 904–916, 2001.

[2] T. Leclère, M. Lavandier, and F. Perrin, "On the externalization of sound sources with headphones without reference to a real source," *J. Acoust. Soc. Am.*, vol. 146, no. 4, pp. 2309–2320, 2019. DOI: 10.1121/1.5128325.

[3] B. Rakerd and W. M. Hartmann, "Localization of sound in rooms, II: The effects of a single reflecting surface," *J. Acoust. Soc. Am.*, vol. 78, no. 2, pp. 524–533, 1985, ISSN: 0001-4966, 1520-8524. DOI: 10.1121/1.392474.

[4] P. Zahorik, D. J. Kistler, and F. L. Wightman, "Sound Localization in Varying Virtual Acoustic Environments," in *2nd International Conference on Auditory Display*, Santa Fe, 1994, pp. 179–186.

[5] R. Guski, "Auditory Localization: Effects of Reflecting Surfaces," *Perception*, vol. 19, no. 6, pp. 819–830, 1990. DOI: 10.1068/p190819.

[6] D. K. Reed and R. C. Maher, "An Investigation of Early Reflection's Effect on Front-Back Localization in Spatial Audio," New York: Audio Engineering Society, 2009, pp. 1–9.

[7] Y. Kamizono, W. Kobayashi, I. Taniguchi, H. Nishikawa, and T. Onoye, "A Platform for Evaluation of Synthetic Reflected Sounds on 3D Sound Localization," *IEEE Access*, vol. 12, pp. 86 906–86 916, 2024. DOI: 10.1109/ACCESS.2024.3415820.

[8] F. Wendt, R. Höldrich, and M. Frank, "The Influence of the Floor Reflection on the Perception of Sound Elevation," in *Fortschritte der Akustik – DAGA 2017*, Kiel: Deutsche Gesellschaft für Akustik (DEGA), 2017, pp. 767–770, ISBN: 978-3-939296-12-6.

[9] M. Væhrens, S. Serafin, F. Christensen, and D. Hammershøi, "Can the individual winner HRTFs be determined in a shooting task during onboarding for an Audio Only VR?" In *160th Audio Engineering Society Convention, Copenhagen*, Copenhagen, Denmark, 2026, p. 8.

[10] M. Væhrens, S. Serafin, F. Christensen, and D. Hammershøi, "Which Tracking Characteristics from an Audio Only VR Onboarding Predict the Best Performance HRTFs?" In *6th International Conference on Audio for Virtual and Augmented Reality and Immersive Games*, 2026, p. 10.

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

[12] J. C. Makous and J. C. Middlebrooks, "Two-dimensional sound localization by human listeners," *J. Acoust. Soc. Am.*, vol. 87, no. 5, pp. 2188–2200, 1990. DOI: 10.1121/1.399186.

[13] V. Tabry, R. J. Zatorre, and P. Voss, "The influence of vision on sound localization abilities in both the horizontal and vertical planes," *Front. Psychol.*, vol. 4, 2013. DOI: 10.3389/fpsyg.2013.00932.

[14] D. Hammershøi and J. Sandvad, "Binaural auralization: Simulating free field conditions by headphones," in *96th Audio Engineering Society Convention (AES), Amsterdam, February 26-March 01, 1994*, ISSN ; - ; Binaural Auralization ; Conf. date: 19-05-2010, 1994, pp. 1–19.

[15] H. E. Heffner and R. S. Heffner, "The sound-localization ability of cats," en, *J. Neurophysiol.*, vol. 94, no. 5, pp. 3653–3655, 2005. DOI: 10.1152/jn.00720.2005.

[16] B. G. Shinn-Cunningham, "The Perceptual Consequences of Creating a Realistic, Reverberant 3-D Audio Display," in *In Proceedings of the ICA 2004*, Kyoto, Japan: The 18th International Congress on Acoustics, 2004.