forum acusticum 2026
8-12 September · Graz, Austria

Comparative Analysis of Sound Localization Accuracy in Real and Virtual Environments Using a Head-Tracked Dynamic Beacon Source

Ivan Ucović[1], Vedran Planinec[1], Marko Horvat[1], Kristian Jambrošić[1]

[1]*Department of Electroacoustics, Faculty of Electrical Engineering and Computing University of Zagreb, Croatia*

Abstract

With the advancement and implementation of Virtual Reality (VR) and Augmented Reality (AR) technologies, there is an increasing demand for improvements in spatial audio that accurately mimics real-world perception. Regarding human auditory perception, sound localization is a crucial element. Previous research has shown that localization accuracy often differs between physical and virtual environments when comparing horizontal (azimuth) and vertical (elevation) planes, with the latter exhibiting more pronounced discrepancies. This study aims to examine the differences in localization accuracy by utilizing a moving sound source (dynamic beacon) rather than traditional static sources. The experimental setup involves participants placed in an acoustically controlled environment surrounded by a loudspeaker configuration, where signals are routed to simulate a continuous trajectory. Participants are instructed to track the moving source with their head, while the movement is measured with a head tracker. This trajectory is subsequently reproduced binaurally using generic Head-Related Transfer Functions (HRTFs), and the tracking task is repeated in the virtual domain. By comparing head-tracking data from both experiments, the study analyzes the precision of binaural synthesis relative to loudspeaker-based Ambisonics reproduction serving as a physical acoustic reference. Preliminary expectations suggest superior localization performance in the loudspeaker-based measurements, particularly within the horizontal plane. The implications and results of these findings will be discussed further, along with the potential uses of individualized HRTFs.

Keywords: *sound localization, head-related transfer function, head tracking, dynamic beacon source*

Corresponding author: *ivan.ucovic@fer.unizg.hr*

1. Introduction

The growing adoption of virtual and augmented reality has increased the demand for spatial audio that reproduces real-world auditory perception convincingly. Sound source localization is central to this, and its accuracy depends on how faithfully a reproduction system preserves the acoustic cues the auditory system relies on.

Horizontal localization is governed primarily by interaural time and level differences, which are reliable cues and relatively easy to reproduce. Vertical localization, in contrast, depends on direction-dependent spectral filtering by the outer ear, and these cues are highly individual [1]. Binaural reproduction over headphones therefore requires Head-Related Transfer Functions (HRTFs), and when generic rather than individually measured sets are used, elevation performance degrades substantially while azimuth remains largely intact [2]. Individualized HRTFs improve localization accuracy [3] but measuring them for every listener is impractical in most applications, so generic sets remain the norm.

Localization is not a static process. Head movements resolve front-back ambiguity and improve accuracy [4], [5], and incorporating head tracking into binaural reproduction is now standard practice [6]. Ambisonics is a convenient format in this context, as the same scene representation can be decoded either binaurally or to an arbitrary loudspeaker layout [7]. Practical implementations, including low-cost head-tracking hardware, are readily available [8], and spectral modifications have been proposed to further reduce front-back confusion in binaural rendering [9].

Most localization studies use static sources and a pointing or verbal response. A smaller body of work has examined how listeners follow sounds that move [10]. One such study measured head tracking of auditory targets in individualized virtual auditory space across a range of velocities, quantifying performance through onset, RMS and gain error [11]. Another showed that listeners follow unpredictably moving free-field sources

with smooth head movements at a consistent delay [12]. Both were restricted to the horizontal plane, and both used either individualized rendering or a physical source.
The present work applies a continuous tracking approach to a reproduction chain built from generic HRTFs and third-order Ambisonics and extends it to the vertical plane. A pink noise source follows a pseudo-random trajectory while the listener tracks it with the head; the reference trajectory and the measured head orientation are recorded in parallel, and tracking performance is quantified by RMS error, temporal lag, and tracking gain. Three conditions are compared: azimuth alone, elevation alone, and both planes simultaneously.
This contribution reports the first phase of a wider study. It presents the measurement system, its validation, and pilot results from a single participant. The loudspeaker-based Ambisonics reference measurement announced in the abstract is currently in preparation and will be reported separately.

2. Methodology

2.1 System Architecture

The measurement system was designed to simultaneously capture head orientation data and reproduce a spatially moving sound source in the binaural domain.
Head orientation was measured using a Supperware head tracker mounted on Audio-Technica ATH-M50x headphones, transmitting quaternion data via the BridgeHead application at a tracking rate of 50 Hz using the Open Sound Control (OSC) protocol [13]. The OSC data stream was distributed by OSCRouter to two destinations: the IEM Scene Rotator plugin (port 7000) for real-time scene compensation, and a MATLAB data acquisition script (port 9000) for recording. The signal flow is illustrated in Fig. 1.
Audio processing was performed in Reaper DAW using the IEM Plugin Suite. The signal chain consisted of: (1) a pink noise generator as the sound source, (2) an IEM Stereo Encoder to position the source within a third-order Ambisonics sound field, (3) an IEM Scene Rotator to compensate for head rotation, and (4) an IEM Binaural Decoder to render the Ambisonics field as a binaural signal for headphone playback. Third-order Ambisonics was selected to provide sufficient spatial resolution for accurate localization, and to maintain consistency with the planned second phase of this study involving a 16-loudspeaker third-order Ambisonics reproduction system. The IEM Binaural Decoder also applied a frequency-dependent compensation filter to correct for the amplitude-frequency characteristics of the Audio-Technica ATH-M50x headphones. The IEM Scene Rotator plugin compensated for the perceptual rotation of the sound scene that occurs during head movement by applying an inverse rotation to the Ambisonics sound field in real time, based on the quaternion data received from the head tracker. Fig. 2 shows the complete plugin chain as configured in the DAW, together with the automation envelopes described in Section 2.2.
Playback was performed through a Focusrite Scarlett 2i2 (4th generation) audio interface at a sample rate of 48 kHz and 24-bit resolution with a buffer size of 256 samples, resulting in an estimated total system latency of approximately 35–40 ms, including audio buffering, plugin processing, and head tracker delay.

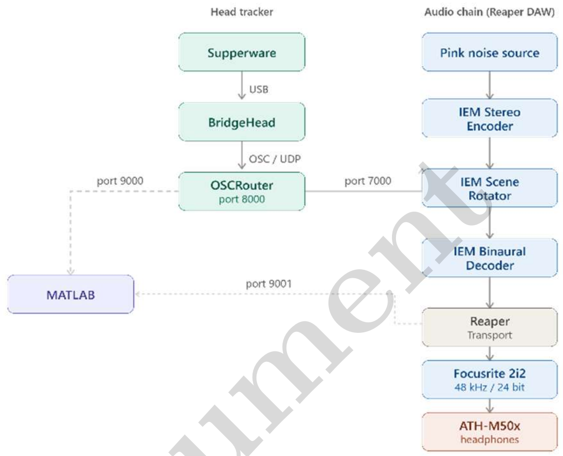

Figure 1. Signal flow of the measurement system, showing head tracking data distribution via OSC and the audio processing chain in the DAW.

Figure 2. Plugin chain and automation envelopes in the DAW, showing the IEM StereoEncoder, SceneRotator with incoming quaternion data, and BinauralDecoder with headphone equalization.

2.2 Sound Source and Trajectory

Pink noise was selected as the test dynamic beacon source due to its broadband spectral content, providing uniform excitation across the audible frequency range. The spatial trajectory of the sound source was defined using automation envelopes for the azimuth and elevation parameters of the IEM Stereo Encoder, programmed as a pseudo-random smooth curve to prevent the participant from anticipating the motion.
The trajectory range was set to ±45° for both azimuth and elevation. The azimuth range was limited to avoid requiring torso rotation, as larger head turns caused unintended shoulder and body movement given the head tracker mounting position on the headphones.

The elevation range was similarly constrained due to the physiological limits of vertical head movement. The total trajectory duration was 28 seconds, chosen to minimize the effect of gyroscope drift accumulation in the head tracker. The resulting trajectory had a mean angular velocity of 12.0 °/s in azimuth and 10.7 °/s in elevation, with peak values of 25.8 °/s and 49.7 °/s respectively. Head tracker drift was not quantified in this study and its influence on short-duration measurements remains to be investigated in future work.

2.3 Measurement Procedure

Three separate measurements were conducted:

Azimuth only - trajectory applied to azimuth; elevation fixed at 0°.

Elevation only - trajectory applied to elevation; azimuth fixed at 0°.

Combined - simultaneous pseudo-random trajectory in both planes.

The participant stood in an upright position, facing forward, which was defined as the zero reference for both azimuth (0°) and elevation (0°). Only head rotation was permitted; torso, shoulder, and body movement was to be avoided. The head tracker was recalibrated before each measurement with the participant in the reference position. MATLAB automatically triggered data recording upon receiving the OSC /play transport message from Reaper and stopped after 28 seconds. Head orientation quaternions were recorded and subsequently converted to Euler angles (yaw, pitch, roll) for analysis.

2.4 Analysis Metrics

Three metrics were used to quantify tracking performance, and all analysis was performed in MATLAB.

The root-mean-square tracking error quantifies the overall angular deviation between the reference trajectory and the measured head orientation:

$$\varepsilon_{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\theta_{ref}(t_i) - \theta_{head}(t_i)\right]^2} \qquad (1)$$

The temporal response lag was determined from the normalized cross-correlation between the two signals, corresponding to the time shift at which their correlation reaches its maximum:

$$\tau = \arg\max_{\tau} R_{\theta_{head}\theta_{ref}}(\tau) \qquad (2)$$

The tracking gain was obtained by linear regression between the reference trajectory and the lag-compensated head orientation:

$$\theta_{head}(t+\tau) = g \cdot \theta_{ref}(t) + b \qquad (3)$$

A gain of unity indicates that the full angular extent of the reference trajectory was reproduced by the head movement, whereas values below unity indicate a systematic reduction of the tracked angular range.

The reference trajectory was exported from the DAW automation envelope and resampled together with the recorded head orientation onto a common time grid of 2000 points spanning the trajectory duration, corresponding to a resolution of 13 ms. The two recordings share a common time origin defined by the OSC transport message that triggers data logging. Lag estimates are therefore quantized to the grid resolution.

3. Results

3.1 Azimuth Tracking

Azimuth tracking yielded the best overall performance among the three conditions. As shown in Fig. 3, the head movement trajectory consistently followed the direction of the reference path, with an RMS tracking error of 14.23°. A systematic undershoot was observed, particularly at the extreme left and right positions, where the participant did not fully reach the peak angular values of the reference trajectory. Tracking accuracy was highest near the center (0°), where directional changes were most perceptually distinct. Cross-correlation analysis (Fig. 6) revealed a peak correlation coefficient of r = 0.92 and a temporal lag of 728 ms, indicating a consistent response delay attributable to perceptual and motor reaction time.

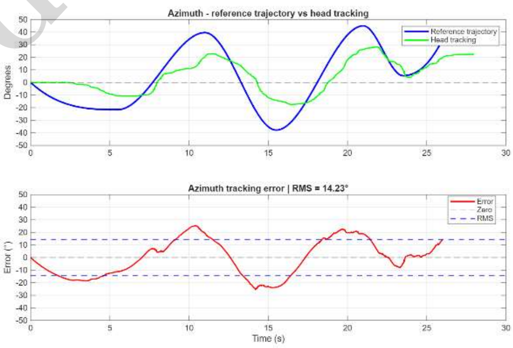

Figure 3. Azimuth tracking in azimuth-only condition. Top: reference trajectory and measured head orientation. Bottom: instantaneous tracking error with RMS bounds.

3.2 Elevation Tracking

Elevation tracking proved significantly more challenging, with an RMS error of 30.98°, more than twice that observed for azimuth (Fig. 4). The participant exhibited a pronounced tendency to perceive the sound source as stationary near the center (0°), even when the reference trajectory reached ±45°. Undershoot was even more pronounced than in the azimuth condition, and the participant showed difficulty distinguishing upward from downward sound motion. Cross-correlation analysis was not performed for elevation, as the participant did not meaningfully track the reference trajectory, making the lag estimate unreliable and difficult to interpret. This behavior is consistent with known perceptual limitations of generic HRTF-based elevation rendering, where the absence of individualized pinnae-related spectral cues reduces the perceptual clarity of vertical spatial information.

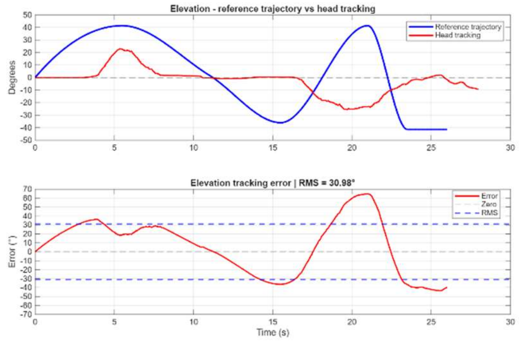

Figure 4. Elevation tracking in elevation-only condition. Top: reference trajectory and measured head orientation. Bottom: instantaneous tracking error with RMS bounds.

3.3 Combined Tracking

In the combined condition, azimuth tracking remained comparable to the azimuth-only measurement, with an RMS error of 11.42°, and a reduced temporal lag of 351 ms (Fig. 6). However, elevation tracking was almost entirely absent, head movements in the vertical plane remained near 0° throughout the measurement, as clearly visible in Fig. 5. This indicates that when both planes are simultaneously active, the participant prioritizes azimuth, while elevation changes elicit little or no head movement. Notably, the RMS elevation error in the combined condition (27.35°) was lower than in the elevation-only condition (30.98°). This counterintuitive result does not indicate improved elevation tracking but rather reflects the near-complete absence of vertical head movement. The participant remained near 0° elevation throughout, which statistically reduces the RMS error when the reference trajectory is symmetrically distributed around 0°.

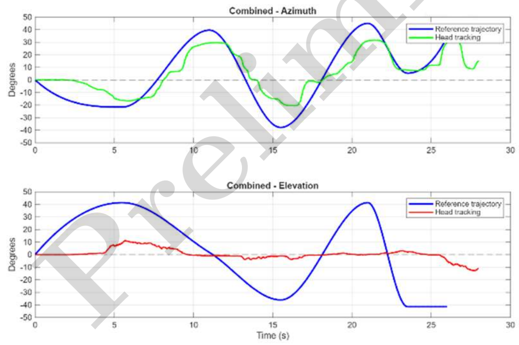

Figure 5. Azimuth and elevation tracking in the combined condition.

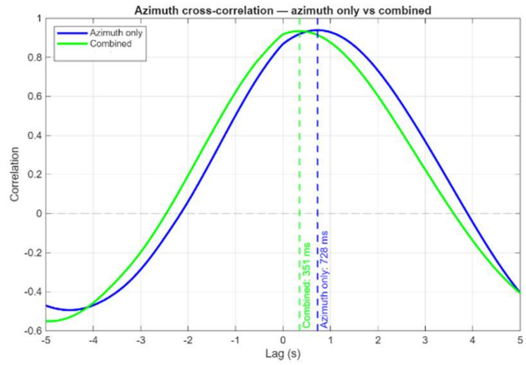

Figure 6. Normalized cross-correlation between reference trajectory and measured head orientation for the azimuth-only and combined conditions, with the corresponding temporal lags indicated.

3.4 Two-Dimensional Trajectory Analysis

Fig. 7 presents the two-dimensional spatial trajectory of the combined condition, with color encoding time progression from blue (0 s) to red (28 s). While the reference trajectory spans a wide area of the azimuth-elevation space, the head tracking data remains concentrated in a narrow region close to the horizontal axis, confirming the dominance of azimuth tracking and the near-complete absence of elevation tracking. The horizontal spread of the head tracking data also reflects the undershoot effect, where the participant consistently tracked the correct direction but with reduced angular range. Table 1 summarizes the tracking metrics across all three conditions; lag and gain are reported for azimuth only.

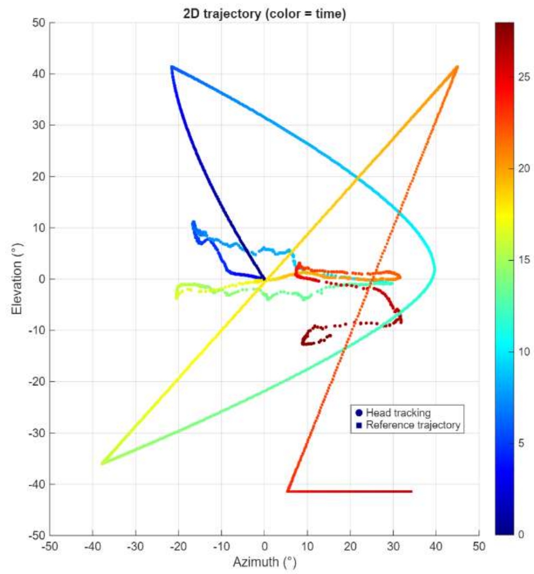

Figure 7. Two-dimensional trajectory of the reference path (continuous line) and measured head orientation (markers) in the combined condition. Color encodes time progression over the measurement duration.

Table 1. RMS tracking error, RMS compensated, gain, and temporal lag across measurement conditions.

Condition	$\boldsymbol{\varepsilon_{RMS}}$ (°)	$\boldsymbol{\varepsilon_{RMS,comp}}$ (°)	$\boldsymbol{g}$	$\boldsymbol{\tau}$ (ms)
Azimuth only	14.23	11.76	0.53	728
Elevation only	30.98	—	—	—
Combined (azimuth)	11.42	10.27	0.62	351
Combined (elevation)	27.35	—	—	—

4. Discussion

The results confirm the expected asymmetry between horizontal and vertical sound localization performance. Azimuth tracking was significantly more accurate than elevation tracking, which is consistent with the underlying psychoacoustic mechanisms. Azimuth perception relies primarily on interaural time differences (ITDs), which are robust and well-preserved even with generic HRTFs, whereas elevation perception depends on spectral cues introduced by the shape of the outer ear (pinnae). Since individualized HRTFs were not used in this study, the elevation cues provided by the IEM Binaural Decoder did not match the participant's own pinnae characteristics, leading to degraded vertical localization.

Undershoot was observed in all conditions: the participant tracked the direction of motion correctly but did not reach the full angular extent of the reference trajectory. Response delay contributes to this, since the source begins to reverse before the head reaches the corresponding extreme. However, the azimuth trajectory reversed direction on average every 5.2 s, roughly seven times the measured lag of 728 ms, so this mechanism alone cannot account for the observed reduction in range. Compensating for the measured response delay reduced the RMS tracking error only modestly, from 14.23° to 11.76° for azimuth-only and from 11.42° to 10.27° for the combined condition, and the regression gain remained well below unity in both cases (0.53 and 0.62). The undershoot is therefore predominantly an amplitude effect rather than a timing effect. Head inertia and neuromuscular dynamics contribute, as the head cannot reverse direction instantaneously, and a genuine compression of the perceived spatial range cannot be excluded from the present data. Separating these contributions requires trajectories at several controlled velocities, which are planned for future work. The estimated system latency of 35–40 ms is negligible in comparison and does not contribute meaningfully to the observed lag. It is worth noting that a lower audio buffer size of 32 samples was tested during system development, which resulted in frequent DAW freezes due to the high volume of OSC messages requiring processing within a very short time window. A buffer size of 256 samples was therefore selected as a stable compromise.

In the combined condition, elevation tracking was almost entirely absent, while azimuth tracking remained comparable to the azimuth-only condition. This is likely explained by attentional dominance, since azimuth is more perceptually accessible, the participant naturally focused on horizontal tracking and disregarded elevation, which was already perceptually ambiguous due to the generic HRTF. Cross-correlation analysis showed a shorter temporal lag in the combined condition (351 ms) compared to the azimuth-only condition (728 ms). This difference should be interpreted with caution at this stage, as it may reflect trajectory familiarity, given that the participant had already heard a similar path in previous measurements, rather than a genuine improvement in tracking response. This warrants further investigation with a larger participant group.

Notably, the RMS elevation error in the combined condition (27.35°) was lower than in the elevation-only condition (30.98°). As discussed in Section 3.3, this is not indicative of better tracking performance but rather reflects the near-complete absence of vertical head movement. In the elevation-only condition, the participant occasionally moved in the wrong direction, confusing upward and downward motion, which increased the RMS error. Repeated neck flexion and extension to ±45° over the full measurement duration approaches the limits of the comfortable range of head movement, so part of the elevation undershoot may also be biomechanical in origin. Additionally, the trajectory speed may have been too fast for reliable elevation tracking, and a slower trajectory is expected to yield better results in future measurements. The azimuth and elevation envelopes were drawn independently and deliberately differed in phase and velocity profile, so that neither plane could be anticipated from the other. A consequence is that the two planes were not matched in velocity, with elevation reaching higher peak rates than azimuth. The difference in tracking performance between the two planes therefore reflects a combination of spatial plane and movement velocity rather than spatial plane alone, although the near-complete absence of elevation tracking in the combined condition, where both planes were active simultaneously, suggests that velocity is not the dominant factor.

This pilot study has several limitations that should be acknowledged. Measurements were conducted with a single participant, making it impossible to draw statistically significant conclusions. The results are highly dependent on how closely the generic HRTF set matches the individual participant's ear geometry. The three conditions were presented in a fixed order without randomization or counterbalancing, so order effects and familiarity with the trajectory cannot be separated from the reported differences between conditions. The planned loudspeaker-based Ambisonics reproduction was not included in this pilot study. Head tracker drift was not quantitatively assessed in this study. The quaternion norm remained within 0.995 to 1.005 throughout all measurements, which indicates that significant drift is unlikely, but a norm close to 1.0 is a necessary rather than sufficient condition for drift-free operation and cannot be considered conclusive evidence of its absence. A dedicated drift assessment is therefore planned for future work. Future work should also include a larger number of participants, individualized HRTF measurements, and a comparison with physical Ambisonics reproduction on a 16-loudspeaker third-order system. It would also be valuable to investigate the effect of trajectory velocity and the rate of direction reversals independently, since a trajectory may be slow yet reverse frequently. Both parameters are likely to have a considerable influence on tracking accuracy.

5. Conclusions

This study presented and validated a real-time measurement system for evaluating head-tracking accuracy in binaural sound source localization using a moving beacon source. Unlike traditional static localization tasks, the proposed methodology allows continuous tracking behavior to be quantified through synchronized head orientation and reference trajectory data. The pilot results demonstrate that the system reliably captures azimuth tracking performance, while confirming the well-established perceptual limitations of generic HRTF-based elevation rendering. These findings, although based on a single participant, establish a working methodology and a set of quantitative metrics, including RMS errors, temporal lag, and tracking gain, which can be directly applied to larger-scale studies. Future work will extend this methodology to individualized HRTFs, multiple participants, varying trajectory speeds and directional dynamics, and a comparison with physical loudspeaker-based Ambisonics reproduction.

References

[1] J. Blauert, Spatial Hearing: The Psychophysics of Human Sound Localization, rev. ed. Cambridge, MA: MIT Press, 1997.

[2] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am., vol. 94, no. 1, pp. 111–123, 1993.

[3] V. Planinec, J. Reijniers, M. Horvat, H. Peremans, and K. Jambrošić, "The accuracy of dynamic sound source localization and recognition ability of individual head-related transfer functions in binaural audio systems with head tracking," Applied Sciences, vol. 13, no. 9, p. 5254, 2023.

[4] H. Wallach, "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol., vol. 27, no. 4, pp. 339–368, 1940.

[5] F. L. Wightman and D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am., vol. 105, no. 5, pp. 2841–2853, 1999.

[6] K. I. McAnally and R. L. Martin, "Sound localization with head movement: Implications for 3-D audio displays," Front. Neurosci., vol. 8, p. 210, 2014.

[7] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," J. Audio Eng. Soc., vol. 60, no. 10, pp. 807–820, 2012.

[8] M. Romanov, P. Berghold, D. Rudrich, M. Zaunschirm, M. Frank, and F. Zotter, "Implementation and evaluation of a low-cost head-tracker for binaural synthesis," presented at the 142nd AES Convention, Berlin, Germany, May 2017.

[9] M. Frank and F. Zotter, "Simple reduction of front-back confusion in static binaural rendering," in Proc. Fortschritte der Akustik – DAGA 2018, Munich, Germany, 2018.

[10] S. Carlile and J. Leung, "The perception of auditory motion," Trends Hear., vol. 20, pp. 1–19, 2016.

[11] J. Leung, V. Wei, M. Burgess, and S. Carlile, "Head tracking of auditory, visual, and audio-visual targets," Front. Neurosci., vol. 9, p. 493, 2016.

[12] J. A. García-Uceda Calvo, M. M. van Wanrooij, and A. J. Van Opstal, "Adaptive response behavior in the pursuit of unpredictably moving sounds," eNeuro, vol. 8, no. 3, ENEURO.0556-20.2021, 2021.

[13] M. Wright, "Open Sound Control: An enabling technology for musical networking," Organised Sound, vol. 10, no. 3, pp. 193–200, 2005.