forum acusticum 2026
8-12 September · Graz, Austria

TOWARDS SPATIAL AUDITORY ATTENTION DETECTION USING DRY ELECTRODE ELECTROMYOGRAPHY AND EYE TRACKING

Mira Hesse[1], Max Stelzenmüller[2,1,*], Maria Hutschenreuter[3], Stephan Werner[1], Patrique Fiedler[3], Andreas Nürnberger[2]

[1] *Electronic Media Technology Group, Technische Universität Ilmenau, Germany*
[2] *Data & Knowledge Engineering Group, Otto-von-Guericke-Universität Magdeburg, Germany*
[3] *Data Analysis in Life Sciences Group, Technische Universität Ilmenau, Germany*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

This study investigates the potential of decoding spatial auditory attention using biosignals, specifically the vestigial postauricular muscle reflex (PAMR) and eye movements. Surface electromyography (EMG) using dry electrodes and eyetracking were recorded concurrently while participants were exposed to acoustic stimulation (broadband noise and narrowband sine bursts) across nine spatial locations and four sensation levels. Results reveal a functional difference between the two systems. The PAMR acts as a rapid, orienting mechanism that requires high-intensity, broadband acoustic energy to trigger a robust spatial gradient, actively suppressing frontal and narrowband targets. The eye movements exhibit a lateralized orienting response that is more robust across all tested spectral conditions.

Keywords: *spatial auditory attention, electromyography, eye tracking, pamr*

1. Introduction

The human auricular muscles possess a hidden but consistently measurable reflex system that responds to acoustic stimulation [1]. A prominent example is the Posterior Auricular Muscle Reflex (PAMR), an electromyographic signal that has been extensively investigated under various basic conditions [1], [2]. Building on the established reliability of the PAMR, this study is motivated by the hypothesis that auricular muscle activity holds potential for the detailed investigation of spatial hearing and auditory attention.

Corresponding author: *max.stelzenmueller@ovgu.de.*

Previous work has shown amplitude differences between ipsilateral and contralateral stimulation for transient stimuli [1]. Most studies relied on simple click stimuli to induce the reflex [1], [2], although comparisons have shown that broadband stimuli yield larger PAMR amplitudes than narrowband signals [3]. More recent work by Strauss et al. [4] and Schroeer et al. [5] have shown evidence of muscle activation in directional active listening environments with more complex acoustic stimuli, such as speech or environmental sounds. However, the behavior of the PAMR needs to be studied in more finely sampled, multi-directional spatial settings. Furthermore, it remains unclear how the continuous interplay of stimulus parameters, such as rise time, spectral properties, and sound pressure level, affects these reflexes.
To address these gaps, this study investigates auricular muscle activity in a spatial listening environment to assess its potential in attention detection. Specifically, the electromyographic activity of three auricular muscles (Posterior Auricular Muscle (PAM), Superior Auricular Muscle (SAM), and Anterior Auricular Muscle (AAM)) was recorded concurrently with eye tracking data. These responses were elicited using auditory stimuli with varying spectral characteristics, rise times, and sound pressure levels to systematically map the spatial and acoustic sensitivity of the auricular reflexes.

2. Methods

This section outlines on the stimuli, the spatial loudspeaker setup, the participants, the recording procedures, and the data post-processing.

2.1. Participants

The 17 participants (6 female, 11 male) aged 20 to 36 years (26.29 years ± 3.95 years), three of whom had prior experience in spatial listening experiments.

2.2. Stimuli

Three different trains of bursts (three bursts in each train) were chosen as stimuli for the experiment. The bursts had the following attributes: White noise bursts (75 ms burst, 75 ms silence, with 100 µs half Hanning window rise and fall time), white noise bursts (75 ms burst, 75 ms silence, with 37.5 ms half Hanning window rise and fall time) and 2 kHz pure-tone bursts (75 ms burst, 75 ms silence, with 100 µs half Hanning window rise and fall time), referenced as WS, WL and S respectively.

Stimuli were bandpass-filtered with cut-off frequencies of 50 Hz and 16 kHz using a 6th order Butterworth filter.

A rise time of 100 µs is consistent with studies investigating the PAMR that use 100 µs clicks as stimuli, finding a strong PAMR such as [1]. Stimulus WS was selected to investigate the influence of a longer rise time on the response. Stimulus S is expected to elicit a PAMR [1], while being more ambiguous in localization [6]. The white noise burst stimuli were chosen for their broadband characteristic which shows good localization accuracy [7].

The trains were presented at random varying Inter-Stimulus Intervals (ISIs) between 2.0 s and 4.0 s. The lower ISI threshold was chosen as a pretest suggested possible eye activity after the stimulus of up to 1.5 s. To avoid fatigue the experiment length was set to 30 min [8]. Each stimulus combination (sound sensation level, stimulus type, speaker direction) was presented 4 times. After every 5 min the experiment was paused for 1 min.

All stimuli were calibrated at the listener's head position using an unweighted equivalent continuous sound level L_{Zeq}. The individual sensation level of each participant for each stimulus was determined following the procedure described in [9] (stepsize 5 dB). The stimuli were presented from the back speaker during this test. The stimuli were then presented at four different levels during the main test, starting at 0 dBSL up to 60 dBSL in steps of 20 dBSL.

The localization accuracy for each stimulus was determined in a pretest by marking the perceived source position on a sheet of paper showing a top-down view of the speaker locations. The pretest results show 97 % ± 5 %, 96 % ± 10 % and 56 % ± 10 % mean localization accuracy across all participants for stimulus WS, WL and stimulus S respectively.

2.3. Spatial loudspeaker setup

Stimuli were presented from nine different loudspeakers, all facing the participant. Eight speakers were placed around the participant 45° apart at a height of 1.30 m with a distance of 2.40 m to the participant and one loudspeaker was placed above the participant at a distance of 1.50 m. The participant was sitting on a chair at the listener position (see Fig. 1). The experiment was set up in the listening lab at TU Ilmenau[1] using Geithain RL906 loudspeakers.

[1] tu-ilmenau.de

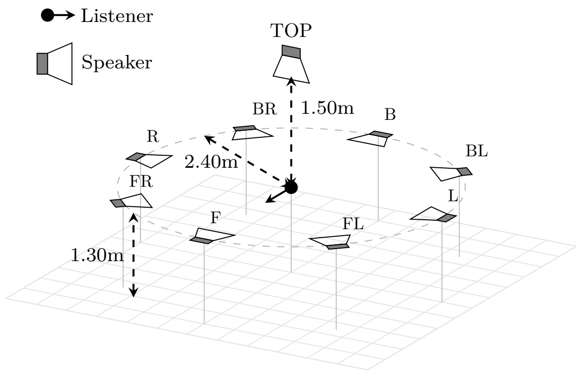

Figure 1: The spatial loudspeaker setup.

2.4. Sensor data

The data recorded during the experiment was collected using two independent sensor systems - a dry electrode Electromyography (EMG) setup and a pair of glasses for eye and head tracking. Whenever amplitude normalization or Z-scoring appears in the results it was performed per participant.

2.4.1. EMG setup and preprocessing

The EMG setup consisted of an eego 24 amplifier with the corresponding recording software and bipolar electrode pairs. The electrodes in each pair were spaced 1 cm apart. Those electrodes were placed on the PAM, the SAM and the AAM on the left and right side of each participant. The reference and the ground electrode were placed on the middle of the forehead and on the left temple respectively. The EMG signals were captured using PU substrate electrodes with silver/silver chloride coating, which were connected to the amplifier via shielded 30 AWG RG178 coaxial cables (50 Ω).

Before preprocessing, for each single electrode of each bipolar pair the signal quality was evaluated for a flat signal (indicating lost contact). These affected signal segments were omitted. Full electrode pairs were omitted when this was true for at least 90 % of the signal. The remaining electrode pairs of the 34 sensors per muscle group (17 participants, bilateral placement), were 11 from the AAM, 20 from the SAM and 34 from the PAM. Therefore the EMG evaluation focused on the PAM data. After this first step the recorded EMG signals were bandpass-filtered with cut-off frequencies of 10 Hz and 300 Hz using a finite impulse response filter. An infinite impulse response (1st order Butterworth) notch filter for 50 Hz and harmonics was applied to remove remaining power-line induced noise. The resulting signals were transformed applying Root Mean Squared (RMS) (7 ms window) and Gaussian smoothing ($\sigma = 7$ ms). From the resulting epochs all those with signal energy higher than 500 µV were excluded. All epochs were baseline corrected to the 200 ms period preceding the first stimulus onset of that epoch. Of the remaining epochs, those with a peak-to-peak amplitude exceeding three times the

standard deviation of the sensor's overall peak-to-peak amplitude were omitted.
Finally, epochs that exceeded the criterion of three times the standard deviation in the eye or head tracking data were also excluded from the EMG dataset.

2.4.2. Eye and head tracking, and preprocessing

To assess eye and head movements during the experiment the Project Aria Gen 1 glasses [10] were used to collect data using the built-in gyroscope and eye tracking cameras. This data was collected for independent analysis, as well as to identify potential artifacts, since eye movements are known to influence auricular EMG recordings [11]. To reduce these potential effects participants were asked to keep facing forward and look at a green dot in the center of the speaker directly in front of them throughout the entire experiment. The head velocity was derived from the vector magnitude of the gyroscope data. The eye tracking cameras were recording at 90 fps and evaluated using the Project Aria eye tracking model[2]. Eye velocity was calculated based on the change in angle over time between consecutive samples.
All Aria data was baseline corrected to the 200 ms period before each trigger, and Gaussian smoothing ($\sigma = 20$ ms) was applied.

3. Results

The recorded data was evaluated based on the continuous time-series epochs as well as the inter-participant variance across the different experimental conditions. To account for repeated measures and isolate the individual effects of the stimulus parameters, linear mixed models (LMMs) were utilized.
For the analysis, EMG results from the left and right sides were grouped together and recategorized based on whether the acoustic stimulus originated from the ipsilateral or contralateral side relative to the recording electrode. Abbreviations follow this scheme: F = front, B = back, S = side, I = ipsilateral, C = contralateral, TOP = top. FI for example references the speaker 45° from the front on the ipsilateral side.

3.1. Epoched time-series data

Mean EMG responses across all participants are shown in Fig. 2 for the PAM. The PAM response exhibits a positive peak in normalized amplitude following each stimulus onset, which then declines below the baseline level after. Coinciding with the offset of each stimulus, a second, smaller positive peak in amplitude is visible. The signal returns to baseline amplitude after approximately 1.0 s.
Fig. 3 illustrates the mean horizontal gaze position (yaw), grouped into left, right, and center stimuli. Center stimuli contain the frontal, top and back speaker, left and right speaker the three speakers on the corresponding side. Eye velocity begins to increase between the first and third stimulus onset and reaches its peak approximately 800 ms after the first stimulus onset, before returning roughly back to baseline level

[2] github.com/facebookresearch/projectAria_eyetracking

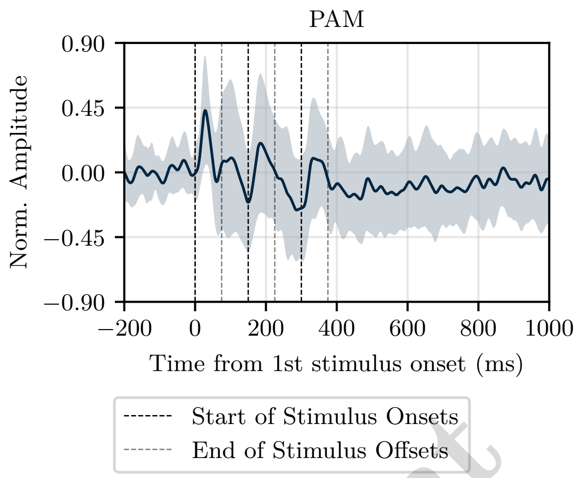

Figure 2: Mean PAM EMG response with standard deviation for all participants.

at 1500 ms. While the velocity profiles are similar across all spatial groups, the eye yaw data reveals clear directional differences. For center stimuli, the mean eye yaw remains centered, with the standard deviation increasing slightly after stimulus onset. For side stimuli, the mean eye yaw exhibits a directional shift towards the side the stimulus was presented on. The standard deviation for these lateralized stimuli is also larger than for the center stimuli.

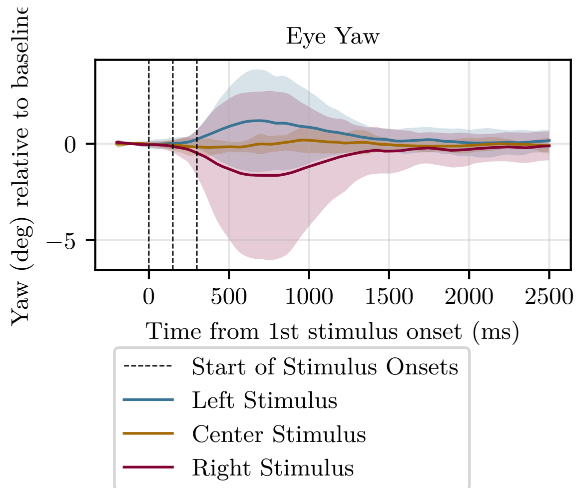

Figure 3: Mean eye responses with standard deviation for all participants.

3.2. LMM results

In all visualizations, plotted values represent model-derived estimated marginal means, with error bars indicating the 95 % confidence intervals.

3.2.1. Interaction between sensation level and speaker direction PAMR

The interaction between speaker direction and sensation level is shown in Fig. 4. The plots show an overall

increase in z-score with sensation level. For the highest sensation levels (40 dBSL and 60 dBSL) the spatial tuning is shows the same pattern across directions. The highest z scores are observed for the ipsilateral side, followed by the contralateral side. Both show an increase from front to back on the corresponding side. The same pattern can be observed for the midline speakers. For the lower sensation levels the pattern is less consistent.
Post-hoc comparisons confirmed that the response to the F speaker was significantly suppressed compared to nearly all other positions in the room, including the B speaker($p < 0.001$), BC speaker($p < 0.001$), and both FI speaker and BI speaker ($p < 0.01$; $p < 0.001$). Conversely, the BI speaker elicited the highest absolute response. This peak response was significantly greater than the response to both the F ($p < 0.001$) and the TOP speaker ($p < 0.01$), as well as the FI speaker ($p < 0.001$) and FC speaker ($p < 0.001$). Notably, while the response clearly lateralized toward the ipsilateral hemisphere, the difference between the BI speaker and SI speaker positions was not statistically significant at 60 dB SL ($p = 0.228$)

3.2.2. Interaction between stimulus type and speaker direction PAMR

The interaction between speaker direction and stimulus type is shown in Fig. 5. For the sine stimulus, spatial tuning is absent for all speaker positions, though small differences can be observed for ipsi- and contralateral stimuli. Post-hoc comparisons confirmed that the this stimulus failed to elicit any significant directional differentiation for all of the nine spatial positions (all $p > 0.35$).
In contrast, both noise stimuli elicited the distinct posterior-ipsilateral pattern as described in Section 3.2.1.

3.2.3. Interaction between sensation level and speaker direction eye yaw

The interaction between speaker direction and sensation level shows no consistent spatial pattern at lower sensation levels (0 dBSL and 20 dBSL). However, at higher sensation levels (40 dBSL and 60 dBSL), a lateralized orienting response shows. Peak eye yaw is positive (indicating leftward movement) for left stimuli and negative (indicating rightward movement) for right stimuli, while midline stimuli elicit near-zero yaw.
Post-hoc comparisons at 60 dBSL confirmed highly significant differences between opposing lateral positions, such as the FL speaker and FR speaker ($p < 0.001$), as well as the L speaker and R speaker ($p < 0.001$). Additionally, midline targets were significantly differentiated from lateral targets, for example between the F speaker and L speaker ($p < 0.001$).
The eye velocity did not show any significant interaction in this context.

3.2.4. Interaction between stimulus type and speaker direction eye yaw

In contrast to the PAMR the interaction between speaker direction and stimulus type for the eye yaw shows spatial tuning across all stimulus types, including the sine stimulus. For the sine stimulus, a lateralized orienting pattern remained evident, and post-hoc comparisons confirmed significant differentiation between opposing lateral targets, such as the FL speaker and FR speaker ($p = 0.010$), and the L speaker and R speaker ($p = 0.001$).
However, noise stimuli (WL and WS) elicited a more precise and statistically robust orienting response. For both broadband types, the lateralization was highly distinct, yielding highly significant contrasts between left and right extreme positions (e.g., FL speaker vs. FR speaker, both $p < 0.001$). For these stimuli a pattern for the left and right stimuli shows, where the eye yaw decreases from front to back (except for the WS stimulus at the left side).

4. Discussion

Descriptive and model based analyses indicate systematic patterns depentent on directional cues, stimulus type and sensation level, for example, increasing EMG amplitude with stimulus sensation level and stimulus dependent differences. However, given the exploratory nature of this study and the limited sample size, these findings should be interpreted as preliminary evidence rather than confirmatory results. Not all electrodes could be placed optimally on every participant, particularly for the AAM and SAM, which contributed to the decision to focus the primary evaluation on the PAM (as detailed in Section 2).

4.1. Influence of sound sensation level

The highest mean z-score of peak to peak amplitude from the PAM was measured at the highest sensation level, which is consistent with earlier studies on the PAMR [1].
However, the data also reveals a greater inter-participant spread at the highest sensation level (60 dBSL). While this could theoretically be attributed to absolute differences in sound pressure level (SPL) between participants, plotting the data on a dB SPL scale confirms that this variance persists independently of volume scale.
Interestingly, responses at 0 dBSL were on average higher than those at 20 dBSL. This may reflect an increase in cognitive load or effortful listening; participants may have actively concentrated on the sound source because they were unsure if a stimulus had been presented. Recent research links vestigial auricular muscle activity to listening effort [5].

4.2. Influence of speaker direction

Looking at the LMM outputs the spatial pattern becomes clearer, but though a trend can be observed, significant differences could mainly be observed to the positions with the strongest (BI) and weakest (F) response. FC, TOP, B and FI positions all show very

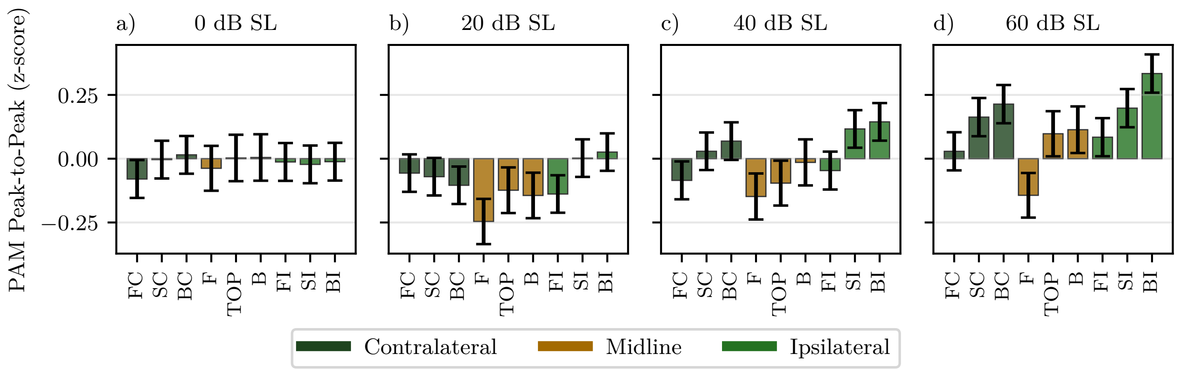

Figure 4: Interaction between Sensation Level and Speaker Direction for the PAM peak-to-peak z-score.

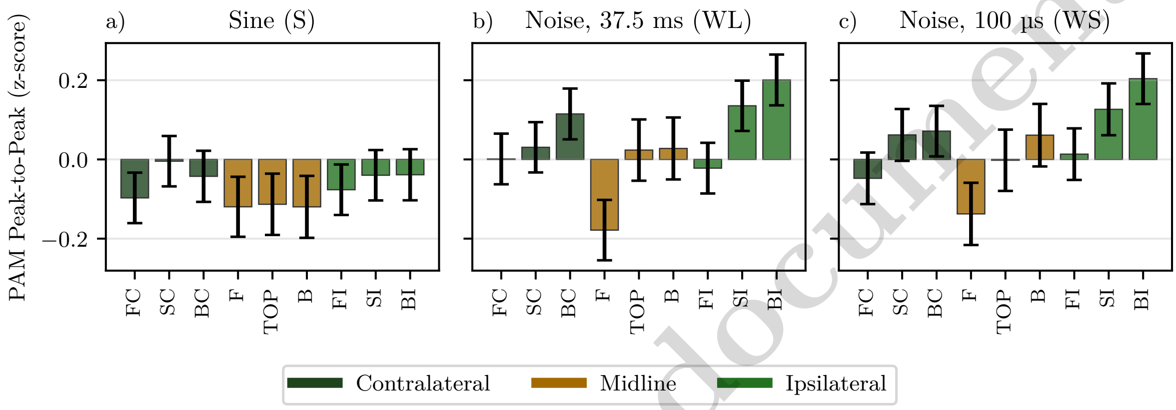

Figure 5: Interaction between Stimulus Type (sine, white noise with 37.5 ms onset and white noise with 100 µs onset) and Speaker Direction for the PAM peak-to-peak z-score.

similar responses indicating challenges for spatial attention detection. For the sine stimulus the spatial pattern almost vanishes, which is consistent with the lower localization accuracy determined in the pretest. This shows especially for the midline stimuli. These stimuli are difficult to differentiate in localization [7], so the results indicate a connection between localization accuracy and spatial tuning in the PAMR. Looking at the eye yaw the spatial tuning for the sine stimulus is still more unclear than for noise stimuli, however a clear left to right difference can be observed and the midline positions behave similarly between all stimuli, indicating that localization ambiguoity is less important for the eye yaw. A possible explanation might be the longer latency of the eye tracking effects.

4.3. Eye movement artifacts

The EMG data exhibits a much higher temporal resolution and a shorter latency in reaction to the trigger than the eye tracking data. Looking at the exact timing, the delayed latency of the initial eye movements indicates that they did not influence the primary EMG peak following the first stimulus onset.

However, the mean eye velocities slowly begin to rise around the second and third stimulus onsets. Because the major oculomotor peak occurs later and at a higher amplitude, a confounding influence on the later EMG peaks cannot be entirely ruled out. How longer stimulus sequences, which extend fully into periods of heightened eye velocity, would influence the EMG remains unexplored. Previous research has shown that active lateral gaze into the stimulus direction significantly increases EMG amplitude [11].

4.4. Real-world applications

The current analysis isolated the pure physiological auricular muscle effects by statistically controlling for confounding variables, such as habituation and individual hearing thresholds, using linear mixed models. This presents a challenge for real-world spatial attention detection, as these confounds may overlap. The high variance in the data between participants and across trials demands a highly robust system architecture. Improving the standard EMG sensor positioning could help mitigate some variability. As the EMG signals offer high temporal resolution while eye tracking provides spatial context, a combination these sensors might help to improve results. Given the amount of interacting conditions, a machine learning solution capable of adapting to complex acoustic environments and individual parameters may be necessary to reliably classify spatial auditory attention in real-time.

5. Conclusion

This study investigated the feasibility of spatial auditory attention detection using biosignals, specifically focusing on the vestigial auricular muscles and eye movement. By concurrently recording surface EMG and eyetracking in a spatial listening environment, the results demonstrate that these sensors offer complementary mechanisms for attention decoding. A functional dissociation between the two systems was found: while the PAMR acts as a fast orienting mechanism dependent on high-intensity broadband acoustic transients, the eye movements exhibit a slower but more universally tuned lateralized orienting response. Despite these promising physiological indicators, deploying such a system in real-world applications faces methodological challenges. The EMG data exhibited substantial inter-participant variance, alongside significant short-term habituation. Future implementations will likely require individual user-calibration phases to optimize spatial tracking performance. As correct electrode placement is important for accurate measurements, high density electrode arrangements are worth investigating, especially when integrating this technology into wearable devices such as headphones. Bridging the gap to real-world applications will require transitioning from controlled, transient stimuli to complex, continuous acoustic environments. Because this study utilized isolated auditory bursts, it primarily captured reflexive, phasic EMG responses. Subsequent research should incorporate continuous real-world signals, such as environmental sounds and speech in multi-speaker scenarios [4], [12], to differentiate the specific triggers for phasic orienting versus sustained tonic auricular muscle activity.

Although auricular muscle activity demonstrates possible feasibility for spatial attention detection, bridging the gap to real-world applications will depend on integrating machine learning models that map muscle responses together with eye tracking data in complex, continuous listening scenarios.

6. Acknowledgements

This work was partially supported by the project "AI Co-Working Laboratories", funded by the European Union through the European Fund for Regional Development (EFRE) as well as the German State of Saxony-Anhalt (Project ID 27813). Collaboration within this project was supported by the Chemnitz-Ilmenau-Magdeburg (CHIM) Research and Innovation Network.

References

[1] G. A. O'Beirne and R. B. Patuzzi, "Basic properties of the sound-evoked post-auricular muscle response (PAMR)," *Hearing Research*, vol. 138, no. 1-2, pp. 115–132, Dec. 1999. DOI: 10.1016/S0378-5955(99)00159-8.

[2] S. C. Purdy, K. B. Agung, D. Hartley, R. B. Patuzzi, and G. A. O'Beirne, "The post-auricular muscle response: An objective electrophysiological method for evaluating hearing sensitivity," *International Journal of Audiology*, vol. 44, no. 11, pp. 625–630, Jan. 2005. DOI: 10.1080/14992020500266639.

[3] R. B. Patuzzi and G. A. O'Beirne, "A correlation method for detecting the sound-evoked post-auricular muscle response (PAMR)," *Hearing Research*, vol. 138, no. 1–2, pp. 147–162, Dec. 1999. DOI: 10.1016/s0378-5955(99)00161-6.

[4] D. J. Strauss, F. I. Corona-Strauss, A. Schroeer, P. Flotho, R. Hannemann, and S. A. Hackley, "Vestigial auriculomotor activity indicates the direction of auditory attention in humans," *eLife*, vol. 9, Jul. 2020. DOI: 10.7554/elife.54536.

[5] A. Schroeer, F. I. Corona-Strauss, R. Hannemann, S. A. Hackley, and D. J. Strauss, "Electromyographic correlates of effortful listening in the vestigial auriculomotor system," *Frontiers in Neuroscience*, vol. 18, Jan. 2025. DOI: 10.3389/fnins.2024.1462507.

[6] W. A. Yost and X. Zhong, "Sound source localization identification accuracy: Bandwidth dependencies," *The Journal of the Acoustical Society of America*, vol. 136, no. 5, pp. 2737–2746, Nov. 2014. DOI: 10.1121/1.4898045.

[7] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. The MIT Press, Oct. 1996, ISBN: 978-0-262-26868-4. DOI: 10.7551/mitpress/6391.001.0001.

[8] N. Pattyn, X. Neyt, D. Henderickx, and E. Soetens, "Psychophysiological investigation of vigilance decrement: Boredom or cognitive fatigue?" *Physiology & Behavior*, vol. 93, no. 1–2, pp. 369–378, Jan. 2008. DOI: 10.1016/j.physbeh.2007.09.016.

[9] R. Carhart and J. F. Jerger, "Preferred Method For Clinical Determination Of Pure-Tone Thresholds," *Journal of Speech and Hearing Disorders*, vol. 24, no. 4, pp. 330–345, Nov. 1959. DOI: 10.1044/jshd.2404.330.

[10] J. Engel et al., *Project Aria: A New Tool for Egocentric Multi-Modal AI Research*, https://arxiv.org/abs/2308.13561v3, Aug. 2023.

[11] R. B. Patuzzi and G. A. O'Beirne, "Effects of eye rotation on the sound-evoked post-auricular muscle response (PAMR)," *Hearing Research*, vol. 138, no. 1–2, pp. 133–146, Dec. 1999. DOI: 10.1016/s0378-5955(99)00160-4.

[12] A. Schroeer, F. I. Corona-Strauss, R. Hannemann, S. A. Hackley, and D. J. Strauss, "The vestigial pinna-orienting system in humans briefly suppresses superior auricular muscle activity during reflexive orienting toward auditory stimuli," *Journal of Neurophysiology*, vol. 132, no. 2, pp. 514–526, Aug. 2024. DOI: 10.1152/jn.00024.2024.