DOI: 10.61782/fa.2025.0218

FORUM ACUSTICUM EURONOISE 2025

SINGLE-TRIAL ELECTROENCEPHALOGRAPHIC ANALYSIS OF FRONT-BACK CONFUSIONS IN FREE-FIELD AND HEADPHONE-BASED LISTENING CONDITIONS

Nils Marggraf-Turley[1*] **Martha Shiell**[2] **Niels Pontoppidan**[2]
Drew Cappotto[1] **Lorenzo Picinali**[1]
[1] Imperial College London, London, UK
[2] Eriksholm Research Centre, Snekkersten, Denmark

ABSTRACT

Sound source localisation relies on spatial auditory cues, which are described by the head-related transfer function (HRTF). In natural free-field listening, individuals perceive sounds filtered by their own HRTFs. Conversely, binaural (headphones-based) reproduction uses HRTFs to render virtual auditory stimuli. However, in practice, binaural reproduction typically employs non-individual HRTFs, which can impair realism and localisation accuracy, introducing phenomena such as front-back and up-down confusion. Beyond behavioural assessment, perceivable differences under the two rendering conditions can be investigated from a neurophysiological perspective. This study uses electroencephalographic data from an existing localisation study to examine disparities in front-back confusion between free-field and headphone-based (non-individual HRTF) conditions. A multilayer perceptron trained on single-trial event-related potentials classified sound source location pairs symmetric around the interaural axis. Single-trial decoding accuracy correlated with behavioural front-back confusion.

Keywords: *Electroencephalography, HRTF, Front-back confusion*

****Corresponding author***: *n.marggraf-turley23@imperial.ac.uk.*

1. INTRODUCTION

The human auditory system localises sounds using spatial auditory cues such as interaural time differences (ITD), interaural level differences (ILD), and monaural spectral cues. These cues, unique to each individual due to anatomical variability, are encapsulated within Head-Related Transfer Functions (HRTFs). Although individual HRTFs enable precise spatial auditory perception, their practical application is limited due to the complexities involved in measurement [1]. Consequently, non-individual HRTFs, often recorded using standardised mannequins such as KEMAR, are widely adopted despite potential inaccuracies, including increased rates of front-back confusion due to inadequate simulation of spectral cues.

Traditional behavioural methods assessing localisation performance with HRTFs rely heavily on subjective reports, which can have low repeatability [2]. Electroencephalography (EEG) provides an objective alternative by recording cortical neural responses associated with auditory spatial perception, potentially bypassing some limitations inherent in subjective feedback.

Previous approaches relating neural response variations to perceptual outcomes show promise [3]. Studies employing event-related potential (ERP)-based decoding have successfully discriminated horizontal and median plane sound locations [4, 5] under free-field listening. More recent work has expanded on these decoding paradigms, indicating lower decoding accuracy for non-individual HRTFs, which was correlated with front-back confusion rates [6]. However, trials were averaged to increase signal-to-noise ratio. Decoding location from a single trial is more advantageous as it paves the way for realtime applications of EEG-based decoding.

Consequently, this study expands on the decoding approach of [6] by using single-trial decoding with a neural network instead of logistic regression with averaged trials.

2. METHOD

We re-examined EEG data from [6], where twenty-two participants localised auditory stimuli under free-field and headphone-based rendering using the KEMAR HRTF from the SONICOM database [7]. Each participant sat in the centre of a spherical loudspeaker array, stabilised with a chin rest. Stimuli comprised a $1000\,\mathrm{ms}$ adapter from 12 azimuthal positions (30° increments), followed by a $100\,\mathrm{ms}$ probe from four locations (Az. 30°, 330°, 150°, and 210°). All stimuli were presented at $65\,\mathrm{dB}$ SPL.

EEG data underwent band-pass filtering (0.1–40 Hz), epoching (-100 to 600 ms) and downsampling (128 Hz). To mitigate artefacts whilst maximising epochs, our artefact rejection differed from [6]. Amplitude thresholding rejected epochs with maximum amplitude above ±200 µV or below ±1 µV. Single-trial ERPs were classified using a multilayer perceptron (MLP) neural network. Feature vectors were constructed by concatenating electrodes and time points, and scaled using Z-score standardisation. The MLP architecture consisted of an input layer, a hidden layer with 512 units (ReLU activation), a dropout layer (0.4), a second hidden layer with 256 units (tanh activation), another dropout layer (0.4), and an output layer with a sigmoid activation function. Training used the Adam optimiser (learning rate of 0.00001) with binary cross-entropy loss. Model performance was evaluated through 10-fold stratified cross-validation for each subject. Data were balanced across subjects, location pairs, and spatial conditions (N=80). Separate networks were trained per subject, location pair (Az. 30 vs 150, Az. 330 vs 210), and spatial condition (Free-field, KEMAR), each for 35 epochs with a batch size of four.

3. RESULTS

Decoding accuracy data were averaged across location pairs and spatial conditions for correlation with behavioural confusion. Front-back confusion rates from behavioural responses were negatively correlated with decoding accuracy ($r = -0.51, p < 0.05$).

4. CONCLUSION

Our findings demonstrate a significant correlation between location decoding accuracy and behavioural front-back confusion rate from single-trial ERPs. This correlation from single-trial decoding highlights the potential for EEG-based approaches to inform HRTF selection or training paradigms aimed at mitigating confusion in headphone-based spatial audio applications.

5. REFERENCES

[1] 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.), Human–Computer Interaction Series, pp. 115–143, Springer, Cham, 2023.

[2] A. Andreopoulou and B. F. G. Katz, "Investigation on subjective HRTF rating repeatability," in *Proceedings of the 140th Audio Engineering Society Convention*, (Paris, France), Paper 9597, Audio Engineering Society, June 2016.

[3] M. G. Wisniewski, G. D. Romigh, S. M. Kenzig, N. Iyer, B. D. Simpson, E. R. Thompson, and C. D. Rothwell, "Enhanced auditory spatial performance using individualized head-related transfer functions: An event-related potential study," *The Journal of the Acoustical Society of America*, vol. 140, pp. EL539–EL544, December 2016.

[4] A. Bednar, F. M. Boland, and E. C. Lalor, "Different spatio-temporal electroencephalography features drive the successful decoding of binaural and monaural cues for sound localization," *European Journal of Neuroscience*, vol. 45, no. 5, pp. 679–689, 2017.

[5] O. Bialas, B. Maess, and M. Schönwiesner, "Evoked responses to localized sounds suggest linear representation of elevation in human auditory cortex," *bioRxiv preprint*, 2023. doi: 10.1101/2023.05.03.539222.

[6] N. Marggraf-Turley, M. Shiell, N. Pontoppidan, D. Cappotto, and L. Picinali, "EEG-Based Decoding of Sound Location: Comparing Free-Field to Headphone-Based Non-Individual HRTFs," 2025. arXiv:2503.10783.

[7] I. Engel, R. Daugintis, T. Vicente, A. O. T. Hogg, J. Pauwels, A. J. Tournier, and L. Picinali, "The SONICOM HRTF Dataset," *Journal of the Audio Engineering Society*, vol. 71, pp. 241–253, May 2023.