DECODING SOUND SOURCE LOCATION FROM EEG: PRELIMINARY COMPARISONS OF SPATIAL RENDERING AND LOCATION

Nils Marggraf-Turley

Dyson School of Design Engineering
Imperial College London
London, UK
`n.marggraf-turley23@imperial.ac.uk`

Lorenzo Picinali

Dyson School of Design Engineering
Imperial College London
London, UK
`l.picinali@imperial.ac.uk`

Niels Pontoppidan, Martha Shiell

Eriksholm Research Centre
Oticon A/S
Snekkersten, Denmark
`npon@eriksholm.com`
`mrhs@eriksholm.com`

Drew Cappotto

Dyson School of Design Engineering
Imperial College London
London, UK
`d.cappotto@imperial.ac.uk`

ABSTRACT

Spatial auditory acuity is contingent on the quality of spatial cues presented during listening. Electroencephalography (EEG) shows promise for finding neural markers of such acuity present in recorded neural activity, potentially mitigating common challenges with behavioural assessment (e.g., sound source localisation tasks). This study presents findings from three preliminary experiments which investigated neural response variations to auditory stimuli under different spatial listening conditions: free-field (loudspeaker-based), individual Head-Related Transfer-Functions (HRTF), and non-individual HRTFs. Three participants, each participating in one experiment, were exposed to auditory stimuli from various spatial locations while neural activity was recorded via EEG. The resultant neural responses underwent a decoding protocol to asses how decoding accuracy varied between stimuli locations over time. Decoding accuracy was highest for free-field auditory stimuli, with significant but lower decoding accuracy between left and right hemisphere locations for individual and non-individual HRTF stimuli. A latency in significant decoding accuracy was observed between listening conditions for locations dominated by spectral cues. Furthermore, findings suggest that decoding accuracy between free-field and non-individual HRTF stimuli may reflect behavioural front-back confusion rates.

1. INTRODUCTION

The ability to localise sounds in space is a fundamental aspect of human auditory perception. This ability not only facilitates the localisation of sound sources but also enables selective attention to specific acoustic targets within complex acoustic scenes [1]. The spatial cues for localisation are well established, with interaural time (ITD) and level (ILD) differences being of primary importance for horizontal lateralisation. Conversely, monaural spectral cues are primarily used for localising elevated sources and discriminating between locations that elicit similar interaural cues [2].

These spatial cues are unique to each listener's Head-Related Transfer Function (HRTF), which represents the frequency-dependent level and time differences of sound propagation at each ear due to interactions with the listener's head, torso and pinnae. HRTFs are widely used to create virtual acoustic environments for Virtual Reality (VR) [3], consumer audio [4] and hearing aid technology [5]. However, generic HRTFs based on average morphologies are often used, which - while effective for simulating interaural cues - are inadequate for replicating spectral cues critical for elevation perception and front-back resolution [7, 6]. Front-back discrimination is also particularly problematic for hearing-impaired listeners who use devices which bypass the majority of their HRTF's spectral cues [8].

Differences in spatial auditory perception between virtual acoustic simulation methods are often evaluated using behavioural measures. These measures rely on the listener's ability to accurately communicate their perception and are as such limited in their efficacy [9]. Moreover, such evaluations have been found to have rather low consistency and repeatability [10, 11], and more in general exclude those unable to engage in traditional behavioural tasks. Advances in noninvasive neuroimaging, particularly the use of electroencephalography (EEG) and magnetoencephalography (MEG), have allowed neural representations of spatial hearing to be explored as alternatives to behavioural measures. The recent integration of EEG into consumer devices like hearing aids [12] and VR headsets [13] presents a particularly exciting opportunity for an objective, non-invasive mechanism for measuring listeners' spatial hearing efficacy without their active involvement, thus mitigating challenges with behavioural measures. Recent work in the field has begun to investigate objective measures of spatial auditory perception focusing on externalisation [14], immersion [15], presence [16], and localisation [17].

While neuroimaging studies have investigated the human auditory system variously under free-field [18, 19], and virtual acoustic cue listening [20, 21], investigations into neural response differences between spatialisation methods are scarce. Palomäki et al. [22] examined variations in the N1m response—an early neural response component—to sound source localisation on the horizontal plane using MEG. Their study assessed the impact of individual HRTFs, non-individual HRTFs, and combinations of ILD and ITD,

as well as isolated cues of each. They found that the amplitude variations of the N1m response in the right hemisphere were most pronounced for individual HRTFs and least for isolated ILD cues. Additionally, the systematicity of these right-hemisphere amplitude changes as the sound source moved from ipsilateral to contralateral directions, was found to correlate with localisation accuracy. This indicates that impoverished spatial cues lead to a less systematic distribution of N1m response amplitudes across locations. Callan et al. [23] identified weaker neural activity associated with internalised stereo listening compared to individual HRTFs using functional magnetic resonance imaging (fMRI). Deng et al. [24] also reported reduced neural activity correlating with the degree of spatial cue impoverishment in a spatial attention paradigm. Here spatialisation methods comprised individual HRTFs, individual ILDs, and generic ITDs. Wisniewski et al. [9] focused on the neural correlates of HRTF individualisation, finding that individual HRTF listening elicited a greater neural response between 300 ms to 500 ms after elevation changes were introduced compared to non-individual HRTFs.

These findings collectively suggest substantial variations in neural responses to varying degrees of spatial cue impoverishment. Such studies use individual HRTFs as their least-impoverished conditions, however, individual HRTFs do not always produce perceptually identical behavioural correlates to free-field listening [25, 26]. This discrepancy suggests the presence of further differences in neural responses to free-field and individual HRTF listening. While animal studies have explored neural responses to free-field versus virtual stimuli [27, 28, 29], comprehensive human studies remain scarce. To the authors' knowledge, only Getzmann et al [30] and [31] include free-field listening in their investigation of spatial cue impoverishment on cortical response to motion onset and distance changes respectively. Both studies found an increase cortical response latency between free-field and virtual listening.

Recently, methodologies for decoding the location of free-field sound sources along the horizontal [18] and median planes [19] have shown promise for evaluating spatial auditory perception. These approaches record Auditory Evoked Potentials (AEP) – the brain's direct electrical response to auditory stimuli [32] – to various sound source locations using EEG, then leverage machine learning models to discern between sound locations based on AEP features. The approach of Bednar et. al. [18] was sensitive to the type of spatial cue (spectral or interaural), with distinct decoding accuracies and response latencies. Here, decoding accuracy (DA) denotes the classification accuracy of the model in distinguishing between sets of sound locations from AEPs. Bialas et al. [19] demonstrated a preliminary link between DA and behavioural localisation performance, suggesting that those with decodable brain responses also exhibited proficiency in sound localisation tasks. This, alongside the above-mentioned findings from Palomäki et al. [22] and Wisniewski et al. [9] that correlate specific neural response components (N1m and P3, respectively) with horizontal and median plane behavioural localisation accuracy, illustrates the potential translation of neural markers into behavioural outcomes.

In summary, the aforementioned studies found that impoverished spatial cues elicit weaker neural activity and also highlight the potential translation of neural responses to perception. The challenge now is to determine the most effective and robust method of interpreting neural responses for the assessment of spatial hearing capabilities. In this paper, we favour a multivariate decoding paradigm established in existing literature [18, 19] to investigate how cortical responses to sound locations vary across different spa-

tial listening conditions. This choice was informed by the increased sensitivity of decoding approaches to differences in the neural encoding of spatial auditory perception that might not become evident when analysing individual components of the neural response [18]. We hypothesise that the degree of spatial cue impoverishment directly influences the accuracy with which sound source locations can be decoded from neural activity. Establishing such a correlation would pave the way for the development of objective, EEG-based localisation metrics, leveraging DA as a measure of efficacy. If such an approach proves consistently effective and repeatable, it may streamline the assessment of spatial hearing and audio rendering techniques.

This paper outlines preliminary findings from three exploratory experiments towards this goal. Given the lack of EEG studies comparing static sound source localisation between free-field listening and virtual listening in humans, our first experiment seeks to compare these conditions. Building upon Bednar et al.'s [18] methodology for free-field decoding, and informed by literature highlighting weaker neural activity with impoverished spatial cues [24, 30, 22, 23], we investigate the feasibility of decoding sound source locations from AEPs in the context of non-individual HRTF listening. The second experiment extends this investigation to DA variations between individual and non-individual HRTF stimuli.

Experiment three employs a modified paradigm from Bialas et al. [19], aiming to improve decodability and achieve decoding between closer sound source locations for both non-individual HRTFs and free-field listening than previously reported [18]. Additionally, the relationship between behavioural front-back discrimination and DA is explored, expanding indications from the free-field median plane decoding study of Bialas et. al. [19] to horizontal plane localisation using non-individual HRTFs.

2. APPARATUS

2.1. Electrophysiology

A single unique subject participated in each experiment. EEG was measured using a BioSemi Active II 64-channel electrode system operating at a sampling rate of 2048Hz. The apparatus adhered to the international 10–20 placement system and was referenced to the average of all electrodes. Data analysis was conducted using the MNE library [33] and custom Python scripts. Data were band-pass filtered between 0.1 to 40Hz, removing slow drifts and maximising signal-to-noise ratio following [14]. Eye movement artefacts were identified using MNE's *find-eog-event* module following [14]. Epochs were then extracted between -100 ms to 600 ms relative to stimulus onset, based on triggers time-locked with the presentation of auditory stimuli. Epochs containing eye-movement artefacts were discarded alongside those where any electrode's peak-to-peak amplitude exceeded $200\,\mu V$ following [30] or was below $1\,\mu V$. To maintain event timing accuracy, data were resampled to 128 Hz after epoch extraction. Baseline correction was applied by subtracting the mean pre-stimulus signal value (-100 ms to 0 ms) from the post-stimulus data.

2.2. Stimulus Delivery System

In all experiments, the participant was seated at the centre of a loudspeaker array, 1.4m from each loudspeaker within a near-anechoic chamber. Head height was adjusted and fixed using a custom chin rest to align ear canals with the horizontal plane of the loudspeaker

< **421** >

array. Auditory stimuli were delivered at 65 dB SPL. These stimuli, generated in Python, were output through an *Antelope Orion 32+* audio interface, with digital-to-analogue conversion at 48kHz. The resulting analogue signals were transmitted either to an array of *Genelec 8010A* loudspeakers or *Audiotechnica ATH-E50* insert earphones for binaural stimuli. Binaural stimuli were synthesized using HRTFs from the SONICOM database and were not head-tracked to avoid potential interference with EEG electrodes [34]. In instances where specific loudspeaker locations did not align with available HRTF data, the 3DTI toolkit [35] applied a barycentric interpolation method among the nearest three HRIRs to generate the required spatial locations. The participant utilised a number pad for varying tasks specific to each experiment. Loudspeaker configurations, stimuli and HRTF specifications for each experiment are detailed in their respective sections.

3. EXPERIMENT 1: DECODING WITH FREE-FIELD VS. NON-INDIVIDUAL HRTF STIMULI

The variation in sound source location DA between free-field and non-individual HRTF listening was investigated. As outlined in Section 1, impoverished spatial cues elicit weaker neural activity, and behavioural localisation accuracy under non-individual HRTF listening is worse for spectral-cue-dominated locations. Accordingly, the following hypotheses were made:

H_1 DA will be lower when decoding between non-individual HRTF stimuli locations than free-field.

H_2 DA disparity will be larger when comparing DA between spectral-cue dominated locations and interaural-cue dominated locations for non-individual HRTF stimuli than for free-field stimuli.

3.1. Experimental Paradigm

The paradigm comprised a single participant being exposed to 100 ms bursts of white noise, played from one of four azimuthal locations: $0°$, $108°$, $180°$, and $288°$, hereafter referred to as Front (F), Right (R), Back (B) and Left (L) respectively. These stimuli were presented with random offset-to-onset intervals between 750 ms to 950 ms to prevent any anticipatory cognitive processes related to the timing of sound onset. For each location and spatial condition, 125 stimuli were presented. The presentation was structured into 20 blocks, with 10 binaural blocks preceding 10 free-field stimuli. Participants were instructed via pre-recorded audio to remove their headphones after completing all binaural blocks. The *KEMAR_GRAS_EarSim_LargeEars_Raw_48kHz* HRTF set from the SONICOM database [34] was used for HRTF processing.

In line with the methodology of Bednar et al. [18], an engagement strategy was employed to maintain participant focus throughout the session. This involved the integration of deviant stimuli, characterised by the concurrent emission of sounds from two distinct locations from the four possible locations. These were presented via loudspeaker or rendered binaurally, depending on the current stimuli block, and had a 10% chance of occurrence. Participants were instructed to acknowledge these deviant stimuli by pressing a key on the number pad as soon as they were perceived. Data from these trials were excluded from the analysis.

3.2. Analysis

A logistic regression classifier was employed following [19] to investigate H_1. The logistic regression model was implemented using the *LogisticRegression* function from the *sklearn.linear_model* module [36]. The classifier was configured to predict between pairwise sound locations based on AEPs observed within a 0 ms to 600 ms time window post-stimulus onset. Data from all 64 electrodes were considered during this window, forming a feature vector of 64 * 77 dimensions per trial. The minimum number of AEPs remaining for any given location after pre-processing and epoching was 105. Therefore, 105 randomly sampled AEPs from each location were used ensuring balanced sets for classification. The model was evaluated using 10-fold cross-validation to provide a robust measure of performance, with stratified folds ensuring a comparable learning problem in each fold.

Listening through headphones using a non-individual HRTF can result in front-back confusion [7]. To this end, AEPs were classified between F-B and L-R locations to investigate H_2. Then, to compare temporal variations of DA across location and spatial condition, a decoder was trained for each time point from -100 ms to 600 ms relative to stimulus onset. Specifically, for each time point, a feature vector of length 64, comprising 1 sample from each electrode, was constructed.

To assess the significance of the DA for both whole window and individual time point decoding, a non-parametric permutation test was conducted, by randomly permuting labels and training the decoder 1000 times to establish a distribution of DAs under the null hypothesis following [18]. The tails of this distribution were then used to calculate significance thresholds. This test was implemented using *sklearn.model_selection.permutation_test_score* [36]. Our analysis differs from [18] in that we use single-trial decoding as opposed to averaging trials. This was done since future applications of assessing localisation efficacy from EEG demand single-trial decoding to not introduce excessive latency between data acquisition and decoding.

Table 1: *Whole-window DA for free-field and non-individual HRTF stimuli. Classification p-value in parentheses.*

	Decoding Accuracy	
Location	Free-Field	KEMAR HRTF
F vs B	65.31% (0.017)	48.67% (0.590)
L vs R	78.02% (0.009)	70.02% (0.010)

3.3. Results

3.3.1. Whole-window decoding

Table 1 summarises the DA for decoding between F-B and L-R locations for free-field and non-individual HRTF stimuli. Significant ($p < 0.05$) decoding was achieved for L-R and F-B location pairs from AEPs elicited by free-field stimuli. Moreover, a lower DA for F-B decoding (65.31%, $p = 0.017$) was observed than for L-R (78.02%, $p = 0.009$). For the non-individual HRTF stimuli, significant decoding was achieved between L-R (70.02%, $p = 0.01$). However, decoding was not possible between F-B locations (48.67%, $p = 0.59$). Finally, the difference in DA between L-R and F-B decoding was larger for free-field than

< **422** >

non-individual HRTF decoding, with differences of 12.71% and 21.35% respectively.

3.3.2. Each time point decoding

Figure 1 shows the variation in DA over time for both spatial conditions and location pairs. DA reaches significance earlier for L-R decoding than for F-B for both free-field and non-individual HRTF decoding. Specifically, for both non-individual HRTF and free-field stimuli, L-R DA peaked at ≈ 80 ms. However, for free-field F-B decoding, clusters of significant decoding are observed between ≈ 100 ms to 380 ms in contrast to ≈ 400 ms to 450 ms for non-individual HRTF stimuli.

3.4. Discussion

Consistent with prior research [18], whole-window DA is significantly above chance for both L-R and F-B locations in free-field conditions, with L-R locations achieving higher DA than F-B locations. Furthermore, F-B locations exhibited a longer latency until DA reached significance when compared to L-R locations, supporting prior findings indicating later processing of spectral cues in the cortex [20, 37]. DA was significant for non-individual HRTF stimuli for L-R locations but lower than free-field decoding, in line with H_1. However, decoding was not possible for F-B locations. Supporting H_2, there was a larger whole-window DA difference between F-B and L-R locations for non-individual HRTF than for free-field stimuli. This raises the question of whether the DA disparity is indicative of increased behavioural front-back confusions. Furthermore, the indication of latency differences between clusters of significant decoding for free-field and non-individual HRTF stimuli for F-B locations, but not for L-R decoding, is intriguing. A possible explanation is that interaural cues are simulated sufficiently by the non-individual HRTF stimuli to not induce cortical response latency. However, spectral cues differ substantially from those that the auditory system is accustomed to, thus causing processing latency. These findings align with similar reports of cortical response latency between real and virtual stimuli motion onset [30] and distance changes [31].

4. EXPERIMENT 2: DECODING WITH FREE-FIELD, INDIVIDUAL AND NON-INDIVIDUAL HRTF

In experiment 1, sound source location was decoded on a single-trial basis between L-R locations – where interaural cues dominate – from non-individual HRTF listening. The indication of larger DA disparity for F-B than L-R locations between free-field and HRTF conditions questions whether DA is reflecting monaural cue efficacy. Thus, Experiment 2 aims to evaluate the DA of sound source locations under individual HRTFs, non-individual HRTFs, and free-field listening. Individual HRTFs simulate spatial cues unique to the listener's morphology with the largest behavioural differences between individual and non-individual processing observed between locations where spectral cues dominate. Thereby we hypothesise:

H_1 DA will decrease from free-field to individual HRTF to non-individual HRTF conditions.

H_2 DA disparity between spatial conditions will be larger for F-B decoding.

4.1. Method

The experimental method was identical to that of Experiment 1, except for the addition of an individual HRTF condition. Individual HRTFs for the participant, identified as P0006 in the SONICOM database, were used [34]. 100 stimuli were played for each location and spatial condition, divided into 10 blocks per spatial condition. The sequence of blocks began with non-individual HRTF presentations, followed by individual HRTF, and concluded with free-field conditions. The analysis procedures were identical to those described in Experiment 1.

4.2. Results

4.2.1. Whole-window decoding

The results, summarised in Table 2, show a decrease in DA from free-field (63%, $p = 0.006$), to individual HRTF (58%, $p = 0.05$), and then to non-individual HRTF (53.2%, $p = 0.25$) for F-B decoding. Free-field and individual HRTF conditions achieve significant DA, whereas decoding F-B locations from AEPs elicited by non-individual HRTF stimuli was not successful. Conversely, all spatial conditions achieved significant DA for L-R locations, with the highest DA observed in free-field stimuli (71.1%, $p = 0.001$), followed by individual (61.3%, $p = 0.016$) and non-individual HRTFs (62%, $p = 0.011$).

4.2.2. Each time point decoding

DA over time varied across spatial conditions and location pairs, as shown in Figure 2. For L-R decoding, significant decoding clusters were identified between ≈ 80 ms to 180 ms across all conditions. For F-B decoding, significant DA was first achieved for free-field stimuli at ≈ 120 ms, with significant clusters observed between ≈ 120 ms to 180 ms. The temporal patterns of DA for individual and non-individual HRTFs differ, with peaks occurring at ≈ 300 ms and ≈ 480 ms respectively. However, qualitatively, no clusters of significant DA were observed for either HRTF condition in F-B decoding with DA peaks disbursed throughout the response time window.

Table 2: *Whole-window DA for free-field, non-individual and individual HRTF stimuli at 2 location pairs. Classification p-value in parentheses.*

Location	Decoding Accuracy		
	Free-Field	Indiv. HRTF	KEMAR HRTF
F vs B	63% (0.006)	58% (0.050)	53.2% (0.250)
L vs R	71.1% (0.001)	61.3% (0.016)	62% (0.011)

4.3. Discussion

The findings corroborate Experiment 1 and related research [18], indicating that DA reaches significance later for F-B than for L-R decoding. Also in line with experiment 1, significant decoding was not possible between F-B locations for non-individual HRTF responses. The dispersal of DA peaks across multiple time points for both HRTF conditions makes the interpretation of latency effects challenging since there seem to be no prominent clusters of significant DA for either condition. Nonetheless, indicating support
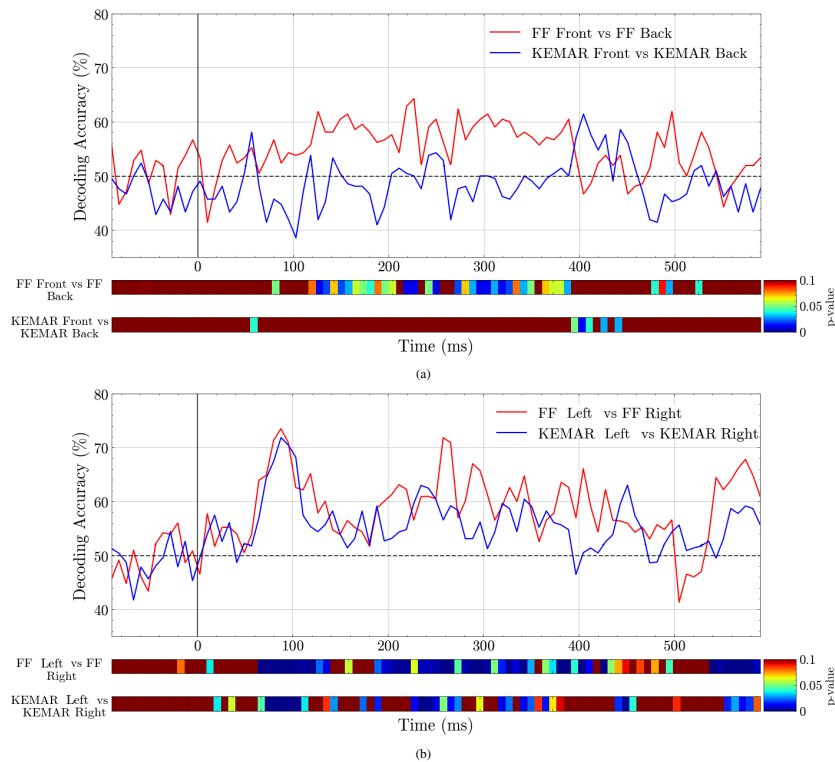
< **423** >

Figure 1: *Exp. 1: Binary classification of F-B versus L-R locations for free-field (FF) and non-individual HRTF (KEMAR) stimuli. (A) Temporal variation of F-B DA. (B) Temporal variation of L-R DA. Colour bars denote DA significance and horizontal lines show chance DA.*

for H_1, whole-window F-B DA decreased from free-field to individual HRTF and to non-individual HRTF listening. Supporting H_2, DA disparity was larger between F-B than L-R for individual and non-individual HRTFs. However, in contrast to H_2, the DA disparity was larger for L-R than for F-B between free-field and HRTF conditions. Participant number prohibits definitive conclusions regarding DA differences, however, these findings indicate that the impoverishment of spectral cues between free-field, individual and non-individual HRTFs may be reflected in the whole-window DA.

5. EXP. 3: BEHAVIOUR AND DECODING ACCURACY

Experiments 1 and 2 suggest differences in DA and DA latency between spatial conditions, calling into question the extent to which cortical response variations correlate with behavioural outcomes. Thereby, experiment 3 investigates the relationship between DA and front-back discrimination under free-field and non-individual HRTF listening. Additionally decoding between smaller spatial resolutions is tested. The following hypotheses are made:

H_1 Behavioural front-back confusion rate will be higher for non-individual HRTF stimuli.

H_2 Higher front-back confusion rates will result in lower DA.

5.1. Method

Our experimental design builds upon the adapter-probe paradigm introduced by Bialas et al. [19]. This approach utilises neural adaptation [38], the reduction in neural activity resulting from repeated or continuous stimulation, to isolate the brain's response to changes in sound location. Specifically, a longer white noise stimulus (the adapter) precedes a short white noise burst (the probe), causing adaptation in sound-responsive neurons. In the instance of adapter-to-probe change, the measured neural response is primarily to changes in location rather than the overall sound onset.

In the the study of Bialas et al. [19], the adapter was played from two free-standing loudspeakers close to the participants' ears resulting in an external probe and internalised adapter. Studies have indicated cortical representations of externalisation [23]. Therefore, in our adaptation, the delivery method of the adapter has been modified to address potential confounds related to changes in externalisation. Specifically, our experiment employs all loudspeakers in the dome for free-field conditions and uses corresponding HRTFs for HRTF conditions, mitigating externalisation changes between adapter and probe. Following [19], our adapter and probe are 1000 ms and 100 ms respectively.

The focus of Experiment 3 was to investigate the relationship between DA and behavioural front-back discrimination. Probe locations comprised two azimuthal locations symmetrical to the interaural axis: $324°$ and $216°$, hereafter referred to as Front-Left (FL) and Back-Left (BL). A 100Hz tone played from the FL position had a 50% chance of occurring after each trial and prompted participants to indicate the perceived location of the last heard probe. Responses were limited to two quadrants FL and BL, associated with two keys on a number pad held by the participants throughout the experiment. ≈ 150 iterations of each location and spatial condition were presented. After pre-processing and separating into epochs, AEPs were categorised into two groups:
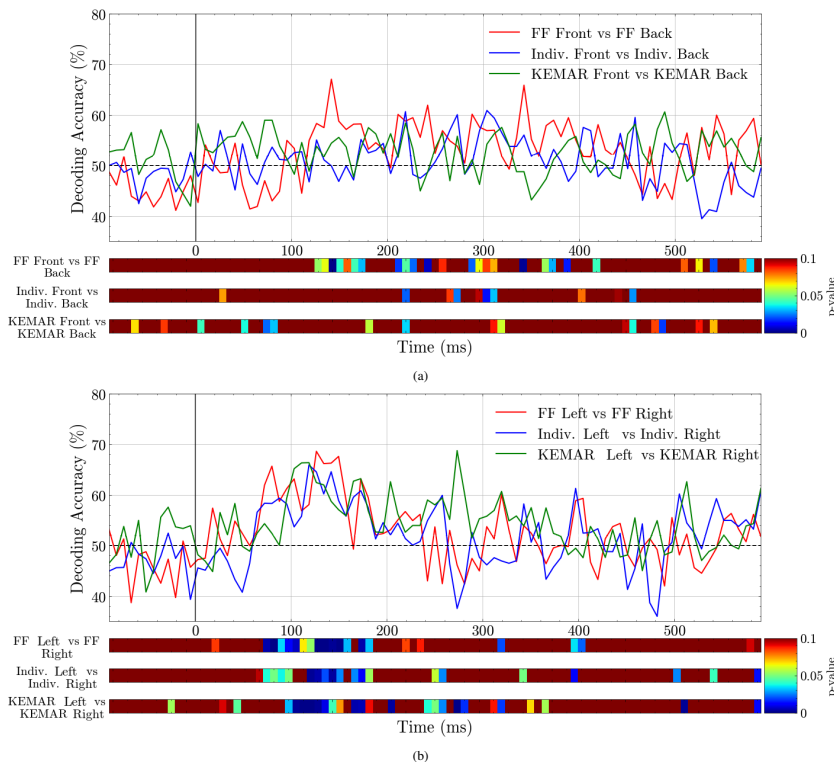
< **424** >

Figure 2: *Exp. 2: Classification of F-B versus L-R locations for free-field (FF), individual and non-individual HRTF (KEMAR) stimuli. (A) Temporal variation of F-B DA. (B) Temporal variation of L-R DA. Colour bars denote DA significance and horizontal lines show chance DA.*

AEPs corresponding to correct behavioural localisation (Correct) and all epochs, irrespective of behavioural response (All), for each spatial condition and location. The minimum number of epochs for any given location and spatial condition in the *All* and *Correct* categories was 126 and 43 respectively. The decoding method was identical to that of experiments 1 and 2.

5.2. Results

5.2.1. Behavioural

Table 3 summarises behavioural results. The participant confused no locations for free-field listening, in contrast to 23% of trials for non-individual HRTF listening. Specifically, 14%, were back-to-front confusions and 9% were front-to-back confusions.

5.2.2. Decoding

When trained on *All* data, the DA for FL-BL was highest for the free-field condition at (63.1%, $p = 0.001$). In contrast, the DA for the Non-Individual HRTF condition was not significant (52.9%, $p = 0.3$). However, when trained on *Correct* responses, significant DA was achieved (62.1%, $p = 0.04$). From 3, it can be seen that DA for free-field stimuli reaches significance earlier than for non-individual HRTF stimuli, with clusters of significant DA observed at ≈ 100 ms and ≈ 345 ms respectively.

5.3. Discussion

The lower number of confusions for free-field listening supports H_1, and lower front-to-back vs back-to-front confusions align with previous studies [7]. Following a similar trend to Experiments 1 and 2, there is a latency in significant DA between non-individual HRTF and free-field stimuli when decoding between locations where spectral cues dominate. Furthermore, whole-window decoding indicates a relationship between DA and front-back confusion errors. Specifically, training the model with a subset of AEPs corresponding to correct behavioural responses improved DA whereas, for free-field stimuli, where no confusions were observed, the DA remains similar across both datasets. These results show initial support for H_2, indicating a relationship between DA and F-B confusion rate. This relationship corroborates with similar indications between free-field median plane localisation and DA [19].

Table 3: *Whole-window DA for free-field and non-individual HRTF stimuli, alongside behavioural F-B confusion rates.*

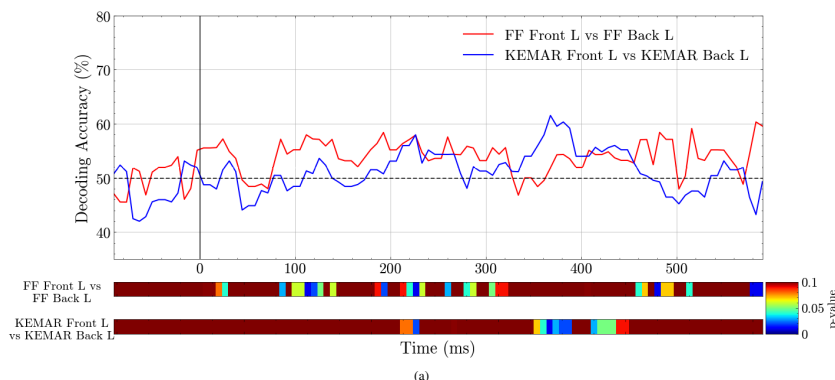|	Free-field	KEMAR HRTF
Confusions	0%	23%
Back-to-front	0%	14%
Front-to-back	0%	9%
DA All (p)	63.1% (0.001)	52.9% (0.300)
DA Correct (p)	62.2% (0.040)	62.1% (0.040)

< **425** >

Figure 3: *Exp. 3: Binary classification of FL-BL locations for free-field (FF) and non-individual HRTF (KEMAR) stimuli. Temporal variation of DA for decoding using all AEPs. Colour bars denote DA significance and horizontal lines show chance DA.*

6. DISCUSSION AND CONCLUSION

Three preliminary studies have been presented, investigating sound source location decoding from AEPs in response to stimuli with varying degrees of spatial cue impoverishment. Each experiment comprised a single unique participant, thereby limiting statistical power which prevented any quantitative assessment of DA differences. Nevertheless, significant decoding for HRTF stimuli was shown, free-field decoding results corroborate previous research and DA differences between conditions suggest new insights for future, more extensive investigations.

Indications from Experiments 1 and 2 are in line with the findings of Bednar et al. [18]. Specifically, significant sound location decoding was achieved for free-field stimuli, and F-B decoding exhibited later DA peaks relative to L-R. In our case, F-B DA reached significance at ≈ 120 ms which is slightly earlier than Bednar et al.'s reported ≈ 150 ms to 200 ms. However, a consistent ≈ 40 ms delay in significant DA between F-B and L-R locations was observed for Experiments 1 and 2, which was similar to their findings. This discrepancy might be attributed to a difference of $\pm 18°$ in the 'Left' and 'Right' positions between studies.

Additionally, these preliminary experiments suggest novel findings, in that significant single-trial L-R decoding with non-individual and individual HRTF stimuli was achieved, with indications of a DA disparity for F-B decoding. This disparity may suggest spectral cue quality is reflected in brain response decodability, aligning with the findings of Bialas et al. [19] linking DA with free-field median plane localisation accuracy.

Experiment 3, adapting Bialas et al.'s approach [19], focused on decoding between locations at lower spatial resolutions with ambiguous interaural cues. Significant DA was achieved only with AEPs of accurate behavioural localisation for non-individual HRTF stimuli, suggesting front-back discrimination ability is reflected in brain response decodability. Latency between clusters of significant DA between non-individual HRTF and free-field stimuli was also observed here. This supports indications from Experiments 1 and 2 that DA latency is observed between spatial conditions for locations where spectral cues dominate, which may suggest cortical processing delay.

While conclusions regarding DA variations are constrained, collectively these studies provide preliminary results indicating the potential of leveraging EEG to evaluate localisation accuracy differences between virtual acoustic simulation methods. To build on these preliminary findings, future research should include larger participant pools and evaluate model efficacy when trained on inter-subject AEPs. Additionally, it would be interesting to explore DA-behaviour relationships in hearing-impaired listeners given their significantly different cortical responses to spatial cue processing [39]. Such studies would determine the extent to which neural response variations and their behavioural associations across different spatialisation methods can be generalised. Establishing repeatable associations would be particularly promising for assessing the efficacy of spatial hearing and spatial rendering methods.

7. REFERENCES

[1] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *J. Acoustical Soc. America*, vol. 25, no. 5, pp. 975–979, 1953.

[2] A. D. Musicant and R. A. Butler, "The influence of pinnae-based spectral cues on sound localization," *J. Acoustical Soc. America*, vol. 75, no. 4, pp. 1195–1200, 1984.

[3] W.-S. Gan, S. Peksi, J. He, R. Ranjan, N. D. Hai, and N K Chaudhary, "Personalized HRTF measurement and 3D Audio Rendering for AR/VR Headsets," *142nd Audio Eng. Soc. Conv.*, 2017.

[4] J.-M. Jot and D. Noh, "Efficient Structures for Virtual Multi-Channel Immersive Audio Rendering," *143rd Audio Eng. Soc. Conv.*, 2017.

[5] R. Eastgate, L. Picinali, H. Patel, M. D'Cruz, "3D Games for Tuning and Learning About Hearing Aids", *The Hearing Journal*, vol. 69, no. 4, pp. 30–32, 2016.

[6] L. Picinali and B.F.G. Katz, "System-to-User and User-to-System Adaptations in Binaural Audio," in M. *Sonic Interactions in Virtual Environments* M. Geronazzo and S. Serafin, Eds., Springer, 2022, pp. 115-143.

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

[8] V. Best, S. Kalluri, S. McLachlan, S. Valentine, B. Edwards, and S. Carlile, "A comparison of CIC and BTE hearing aids for three-dimensional localization of speech," *Int. J. of Audiology*, vol. 49, no. 10, pp. 723–732, 2010.

< **426** >

[9] M. G. Wisniewski, G. D. Romigh, S. M. Kenzig, D. 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," *J. Acoustical Soc. of America*, vol. 140, no. 6, pp. 539–544, 2016.

[10] C. Kim, V. Lim, L. Picinali "Investigation Into Consistency of Subjective and Objective Perceptual Selection of Non-individual Head-Related Transfer Functions" *J. Audio Eng. Soc.*, vol. 68, no. 11, pp. 819–831, 2020.

[11] A. Andreopoulou, B. F. G. Katz, "Investigation on Subjective HRTF Rating Repeatability" *140th Audio Eng. Soc. Conv.* 2016.

[12] C. B. Christensen, R. K. Hietkamp, J. M. Harte, T. Lunner, and P. Kidmose, "Toward EEG-Assisted Hearing Aids: Objective Threshold Estimation Based on Ear-EEG in Subjects With Sensorineural Hearing Loss," *Trends in Hearing*, vol. 22, 2018.

[13] G. Bernal, N. Hidalgo, C. Russomanno, and P. Maes, "Galea: A physiological sensing system for behavioral research in Virtual Environments," in *IEEE Conf. Virtual Reality and 3D User Interfaces*, 2022, pp. 66–76.

[14] T. Colas, N. Farrugia, E. Hendrickx, and M. Paquier, "Sound externalization in dynamic binaural listening: A comparative behavioral and EEG study," *Hearing Research*, vol. 440, 2023.

[15] R. Nicol, O. Dufor, L. Gros, P. Rueff, and N. Farrugia, "EEG measurement of binaural sound immersion," *EAA Spatial Audio Sig. Proc. Symp.* pp. 73–78, 2019.

[16] S. Zhang, X. Feng, and Y. Shen, "Investigating the Potential Use of EEG for the Objective Measurement of Auditory Presence," *Applied Sciences*, vol. 12, no. 5, pp. 2647, 2022.

[17] C. J. Chun, S. H. Jeong, J. W. Shin, H. K. Kim, and J. A. Kang, "Feasibility Study for Objective Measurement on Sound Localization Using Auditory Evoked Potential," in *10th Int. Conf. on Intelligent Inform. Hiding and Multimedia Sig. Proc.*, 2014, pp. 610–613.

[18] 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 J. of Neuroscience*, vol. 45, no. 5, pp. 679–689, 2017.

[19] O. Bialas, B. Maess, and M. Schönwiesner, "Evoked Responses to Localized Sounds Suggest Linear Representation of Elevation in Human Auditory Cortex," 2023.

[20] N. Fujiki, K. A.J. R., V. Jousmäki, J. P. Mäkelä, and R. Hari, "Human cortical representation of virtual auditory space: Differences between sound azimuth and elevation," *European J. of Neuroscience*, vol. 16, no. 11, pp. 2207–2213, 2002.

[21] Y. Wang, Z. Liu, and F. Chen, "An Eeg-Based Study of Auditory Sensitivity in Spatial Hearing," in *Int. Workshop on Acoustic Sig. Enhancement*, 2022, pp. 1–5.

[22] K. J. Palomäki, H. Tiitinen, V. Mäkinen, P. J.C. May, and P. Alku, "Spatial processing in human auditory cortex: The effects of 3D, ITD, and ILD stimulation techniques," *Cognitive Brain Research*, vol. 24, no. 3, pp. 364–379, 2005.

[23] A. Callan, D. E. Callan, and H. Ando, "Neural correlates of sound externalization," *NeuroImage*, vol. 66, pp. 22–27, 2013.

[24] Y. Deng, I. Choi, B. Shinn-Cunningham, and R. Baumgartner, "Impoverished auditory cues limit engagement of brain networks controlling spatial selective attention," *NeuroImage*, vol. 202, pp. 116151, Nov. 2019.

[25] A. W. Bronkhorst, "Localization of real and virtual sound sources," *J. Acoustical Soc. America*, vol. 98, no. 5, pp. 2542–2553, 1995.

[26] F. L. W. and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," *J. Acoustical Soc. America*, vol. 85, no. 2, pp. 868–878, 1989.

[27] O. Behrend, B. Dickson, E. Clarke, C. Jin, and S. Carlile, "Neural Responses to Free Field and Virtual Acoustic Stimulation in the Inferior Colliculus of the Guinea Pig," *J. of Neurophysiology*, vol. 92, no. 5, pp. 3014–3029, 2004.

[28] C. H. Keller, K. Hartung, and T. T. Takahashi, "Head-related transfer functions of the barn owl: Measurement and neural responses," *Hearing Research*, vol. 118, no. 1-2, pp. 13–34, 1998.

[29] R. A. A. Campbell, T. P. Doubell, F. R. Nodal, J. W. H. Schnupp, and D. J. King, "Interaural Timing Cues Do Not Contribute to the Map of Space in the Ferret Superior Colliculus: A Virtual Acoustic Space Study," *J. of Neurophysiology*, vol. 95, no. 1, pp. 242–254, 2006.

[30] S. Getzmann and J. Lewald, "Effects of natural versus artificial spatial cues on electrophysiological correlates of auditory motion," *Hearing Research*, vol. 259, no. 1-2, pp. 44–54, 2010.

[31] B. Stodt, D. Neudek, S. Getzmann, E. Wascher, and R. Martin, "Comparing auditory distance perception in real and virtual environments and the role of the loudness cue: A study based on event-related potentials," *Hearing Research*, vol. 444, pp. 108968, Mar. 2024.

[32] M. P Paulraj, K. Subramaniam, S. B. Yaccob, A. H. B. Adom, and C. R Hema, "Auditory Evoked Potential Response and Hearing Loss: A Review," *Open Biomed Eng J.*, vol. 9, no. 1, pp. 17–24, 2015.

[33] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, 2013.

[34] I. Engel, R. Daugintis, T. Vicente, A. O. T. Hogg, J. Pauwels, A. J. Tournier, and L. Picinali, "The SONICOM HRTF Dataset," *J. Audio Eng. Soc.*, vol. 71, no. 5, pp. 241–253, 2023.

[35] M. Cuevas-Rodríguez, L. Picinali, D. González-Toledo, C. Garre, E. De La Rubia-Cuestas, Luis Molina-Tanco, and Arcadio Reyes-Lecuona, "3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation," *PLOS ONE*, vol. 14, no. 3, pp. e0211899, 2019.

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[37] J. Kaiser, W. Lutzenberger, and N. Birbaumer, "Simultaneous bilateral mismatch response to right- but not leftward sound lateralization:," *NeuroReport*, vol. 11, no. 13, pp. 2889–2892, 2000.

[38] J. Benda, "Neural adaptation," *Current Biology*, vol. 31, no. 3, pp. R110–R116, 2021.

[39] J.-H. Han, J. Lee, and H.-J. Lee, "Ear-Specific Hemispheric Asymmetry in Unilateral Deafness Revealed by Auditory Cortical Activity," *Frontiers in Neuroscience*, vol. 15, pp. 698718, 2021.

<   **427**   >