Where is the cocktail party? Decoding locations of attended and unattended moving sound sources using EEG

Adam Bednar [a,b,*], Edmund C. Lalor [a,b,c,**]

[a] School of Engineering, Trinity College Dublin, Dublin, Ireland
[b] Trinity Center for Bioengineering, Trinity College Dublin, Dublin, Ireland
[c] Department of Biomedical Engineering, Department of Neuroscience, University of Rochester, Rochester, NY, USA

ABSTRACT

Recently, we showed that in a simple acoustic scene with one sound source, auditory cortex tracks the time-varying location of a continuously moving sound. Specifically, we found that both the delta phase and alpha power of the electroencephalogram (EEG) can be used to reconstruct the sound source azimuth. However, in natural settings, we are often presented with a mixture of multiple competing sounds and so we must focus our attention on the relevant source in order to segregate it from the competing sources e.g. 'cocktail party effect'. While many studies have examined this phenomenon in the context of sound envelope tracking by the cortex, it is unclear how we process and utilize spatial information in complex acoustic scenes with multiple sound sources.

To test this, we created an experiment where subjects listened to two concurrent sound stimuli that were moving within the horizontal plane over headphones while we recorded their EEG. Participants were tasked with paying attention to one of the two presented stimuli. The data were analyzed by deriving linear mappings, temporal response functions (TRF), between EEG data and attended as well unattended sound source trajectories. Next, we used these TRFs to reconstruct both trajectories from previously unseen EEG data. In a first experiment we used noise stimuli and included the task involved spatially localizing embedded targets. Then, in a second experiment, we employed speech stimuli and a non-spatial speech comprehension task.

Results showed the trajectory of an attended sound source can be reliably reconstructed from both delta phase and alpha power of EEG even in the presence of distracting stimuli. Moreover, the reconstruction was robust to task and stimulus type. The cortical representation of the unattended source position was below detection level for the noise stimuli, but we observed weak tracking of the unattended source location for the speech stimuli by the delta phase of EEG. In addition, we demonstrated that the trajectory reconstruction method can in principle be used to decode selective attention on a single-trial basis, however, its performance was inferior to envelope-based decoders. These results suggest a possible dissociation of delta phase and alpha power of EEG in the context of sound trajectory tracking. Moreover, the demonstrated ability to localize and determine the attended speaker in complex acoustic environments is particularly relevant for cognitively controlled hearing devices.

1. Introduction

Our auditory system allows us to identify the location of multiple sound sources in our environment and to follow those sources as they change position. This is behaviorally important in and of itself, but also helps us to attend to and understand speech in complex listening scenarios (Cherry, 1953; Shinn-Cunningham et al., 2001). While a substantial body of literature exists on how auditory space is encoded in subcortical neurons (see review by Grothe et al., 2010) less is known

about cortical representation of auditory space. This is particularly relevant for understanding how we operate (pay attention) in complex, realistic scenes. However, despite the ecological relevance of spatial hearing, much work remains to be done to fully characterize how cortex represents of the location of multiple auditory objects in complex scenes.

The vast majority of neurophysiological studies investigating the neural encoding of auditory space have employed a single sound source. At the cortical level, it appears that sound location is processed along the dorsal "where" pathway while object recognition is performed in the

ventral "what" pathway (Romanski et al., 1999; Rauschecker and Tian, 2000; Arnott et al., 2004; Ahveninen et al., 2006a). Human studies have identified that planum temporale and posterior superior temporal gyrus (pSTG) are part of the "where" pathway and respond to spatial sound (Warren et al., 2002; Brunetti et al., 2005; Krumbholz et al., 2005a; Ahveninen et al., 2006b; Deouell et al., 2007; Derey et al., 2016). Electrophysiological studies found that spatially-sensitive neurons have relatively broad spatial receptive fields and the majority of them respond best to sounds from the contralateral hemifield. Nevertheless, other neurons that preferentially respond to central or ipsilateral locations were also observed (Middlebrooks and Pettigrew, 1981; Imig et al., 1990; Stecker et al., 2003; Harrington et al., 2008). Based on this, studies have suggested that the sound source location can be represented using population coding of these relatively broadly tuned neurons (Furukawa et al., 2000; Stecker et al., 2003). More specifically, it has been proposed that the spatial location might be encoded using an "opponent-coding" model, which describes that the sound location can be encoded by relative activation of predominantly contralaterally tuned neural populations. This has been supported by animal (Stecker et al., 2005; Werner-Reiss and Groh, 2008; Ortiz-Rios et al., 2017) as well as human studies (Palomaki et al., 2000, 2005; Krumbholz et al., 2005b; McLaughlin et al., 2015; Derey et al., 2016).

How auditory space is represented in more complex acoustic situations with multiple sources has been less thoroughly investigated. A study by Middlebrooks and Bremen (2013) investigated neurons in cat primary AC when the animal was presented with two spatial sound sequences. The authors found that neurons preferentially synchronized to one of the two stimuli and, in effect, segregated the streams. Also, from human studies on auditory scene analysis, we know that spatial cues alone can facilitate the stream segregation even when other cues are unavailable (Best et al., 2004; Middlebrooks and Onsan, 2012; Carl and Gutschalk, 2013). This suggests we can process overlapping spatial cues corresponding to several sound sources and one can speculate that the spatial representations of multiple sound sources might exist within the ascending auditory pathway. However, it has been proposed that distinct neural pathways are involved in spatial stream segregation vs sound location processing (Middlebrooks and Onsan, 2012) and a recent fMRI study demonstrated that spatial separation between sources but not their locations can be decoded from the auditory cortex (Shiell et al., 2018). Therefore, what exactly happens to neural spatial coding in more complex multi-source scenarios when the auditory system needs to localize several overlapping sounds is unknown.

When investigating hearing in acoustic scenarios with multiple sources, it is important to consider the effects of top-down attentional mechanisms on cortical responses. One of the most common examples of top-down selection in complex acoustic environment is the so called "cocktail party problem", which describes a situation when one is attending to one speaker in the presence of other competing speakers (Cherry, 1953). In such situations, it has been shown that the auditory cortex primarily tracks the dynamic changes of the attended stream. Nevertheless, the unattended stream tracking is also reflected in the cortex although the tracking is weaker (Ding and Simon, 2012; Mesgarani and Chang, 2012; O'Sullivan et al., 2015). With respect to spatial sensitivity to sound, the modulation of cortical responses by attentional task has been shown by single-source studies in animal (Lee and Middlebrooks, 2011) as well as human studies (Ahveninen et al., 2006b). Nevertheless, it appears that attention to the stimuli is not necessary to elicit location-sensitive cortical responses (Deouell et al., 2007). Based on that, one could reason that in multi-source acoustic environments the brain would predominantly represent the position of the attended source and to a lesser extent the position of the other unattended sound sources. However, to our best knowledge no study has addressed the issue of cortical spatial sensitivity to competing sounds in multi-source scenarios.

In this EEG study, we used a stimulus-reconstruction framework from a previous study, which showed that it is possible to reconstruct the sound trajectory of a single sound source (Bednar and Lalor, 2018). Here

we wanted to investigate how cortex encodes spatial information in a multi-source acoustic scenario and we wanted to assess the effect of attention on our ability to decode the locations of simultaneous, competing stimuli. One way to do this is to have both stimuli on continuously and to vary their location over time. As such, in this EEG study, we designed an experiment where listeners attended to one of two simultaneously presented sound stimuli, which were perceived to be randomly moving within the horizontal plane. In particular, the use of stimuli that are moving on independent trajectories allowed us to control (partial-out) for the trajectory of one stimuli and investigate the spatial effects of the other source in isolation.

We tested the trajectory reconstruction in two different experiments that differed in their stimulus and behavior task: Experiment 1 employed continuous band-passed filtered noise stimuli with an active sound localization task. This builds on our previous study where we successfully reconstructed the trajectory of noise stimuli in a single source scenario (Bednar and Lalor, 2018) and we were interested in applying this approach in a multi-source environment. The active localization task was used to motivate subjects to pay attention to the spatial aspect of the auditory scene. In experiment 2, we aimed to test whether the trajectory reconstruction works for more complex speech stimuli in a naturalistic setting, where subjects were given a speech-comprehension task rather than a sound localization task.

In our previous study, we found that we can reconstruct a single sound source trajectory from the phase of delta (<2 Hz) and power of alpha (8–12 Hz) EEG. Here, we predicted that in the multi-source scenario the delta and alpha components of EEG might be differentially sensitive to bottom-up source location and top-down attention: (1) We predicted that the delta EEG component would track the trajectory of both the attended and the unattended sources, however, the unattended would be tracked to a lesser extent. This hypothesis was based on envelope-reconstruction studies, which showed that the low-frequency cortical oscillations predominantly synchronize with the temporal structure of the attended sound stimulus, and less with the unattended source (Ding and Simon, 2012; O'Sullivan et al., 2015). (2) We predicted that the alpha power tracking would be restricted to the attended sound as it has a long-standing association with allocation of spatial attention (Kerlin et al., 2010; Wöstmann et al., 2016).

2. Methods

Data statement. Data and code are available upon direct request.

Participants. In total 29 participants (median = 20 years; min = 18 years; max = 30 years; 12 females; 23 right handed) participated in this study with informed consent. Nineteen subjects took part in the first experiment and ten in the second. Two subjects were excluded from the first experiment due to inability to perform the task. All subjects reported no neurological diseases and normal hearing. The experiments were approved by the Research Subjects Review Board of the University of Rochester.

Experimental Procedures. Participants were presented with two simultaneous continuously moving auditory stimuli via headphones and were asked to pay attention to one of the presented audio stimuli, 'attended' stimulus, while ignoring the other 'unattended' stimulus.

There were two experimental conditions, which differed in the type of auditory stimulus and the task performed by the subjects. In experiment 1, we used filtered noise stimuli and subjects performed a spatial task. In experiment 2, subjects listened to continuous speech stimuli and performed a speech comprehension task.

In both experiments, the attended and unattended stimuli were spatialized to be perceived as randomly moving within the frontal part of horizontal plane using Oculus Audio SDK, which simulates head-related transfer function (HRTF) filtering. The pseudo-random trajectories for both stimuli were generated independently, simulating smooth but unpredictable sound movement. The stimuli were set to move on a semicircular trajectory in the horizontal plane between $-90°$ (left) and

+90° (right) relative to the subject (see Fig. 1A and B). The simulated motion of the source had an average angular velocity of 50°/second. During the experiment, the subjects sat in a dark soundproof room. To minimize movement, the participants were asked to look at a fixation cross displayed on a computer screen directly in front of them.

Experiment 1- Moving Noise Stimuli. Subjects were presented with two moving concurrent band-passed noise stimuli, each having different spectral content, 'Low' and 'High'. Both stimuli were continuous band-pass filtered noise with a bandwidth of the Low stimuli being 250–1250 Hz and 4000–5000 Hz for the High stimuli (see Fig. 1C and D). Before each trial, the subjects were cued to attend to either the Low or High stimulus by text on the computer screen.

The subjects performed a target localization task. The task required subjects to respond with a button press to infrequent tremolo targets (modulation frequency 4 Hz, 2 s long) in the attended stream and indicate whether the target originated from the left or right side of the auditory space using left or right mouse buttons. After each trial, the subjects were shown a target detection score on the screen. The subjects undertook a total of 44 trials, each trial lasting 2 min. The number of targets within each trial ranged from 1 to 4 per trial. The experiment was divided into 4 blocks, each having 11 trials. In experimental blocks 1 and 3 subjects attended to the Low stimulus, while in blocks 2 and 4 attended to the High stimulus.

Experiment 2- Moving Speech Stimuli. Subjects were presented with two different speech streams (audiobooks), each read by different male speaker. Silent gaps in the audio streams exceeding 0.5 s were truncated to 0.5 s in duration. See O'Sullivan et al. (2015) for details. Again, the stimuli were spatialized, so the speakers performed random continuous motion in the horizontal plane. Before each trial, subjects were cued to attend to one of the speakers. See Fig. 1D for example stimuli.

The subjects performed a speech comprehension task and were required to answer between 4 and 6 multiple-choice questions on both stories after each trial. Each question had 4 possible answers. After answering the question, the subjects were shown correct answers. The subjects undertook a total of 40 trials, each trial lasting 1 min. The experiment was divided into 2 blocks, each having 20 trials. For the first 20 trials the subjects paid attention to one speaker and for the next 20 trials they paid attention to the other speaker.

EEG data preprocessing. The EEG data were recorded using a 128-channel ActiveTwo acquisition system (BioSemi, The Netherlands) at a sampling rate of 512 Hz. Preprocessing was done in MATLAB using custom written scripts and the EEGLAB toolbox (Delorme and Makeig, 2004). The data were filtered between 0.02 and 30 Hz and downsampled to 64 Hz. Bad channels were interpolated from the surrounding channels using the spline function from EEGLAB. Independent component analysis

(ICA) was used to remove eye movement artifacts. The components to be removed were preselected by the ADJUST algorithm (Mognon et al., 2011) and then marked for rejection by visual inspection. Finally, the data were re-referenced to the average of all electrodes.

In the first step of our analysis we focused on what EEG frequencies are best for stimulus trajectory reconstruction. We extracted different EEG bands using a sliding band-pass filter from 0.02 to 30 Hz which had a passband window with a width of 2 Hz. We then ran our decoding analysis using each of these EEG bands. We also ran the analysis using the EEG signal power by calculating the analytic envelope of the EEG signal using the Hilbert transformation following the band-pass filtering. The results revealed that the decoding is driven by the lowest frequency EEG delta band (0.02–2 Hz) and by the EEG power at the alpha band (8–12 Hz). In further analysis, we therefore focused on these two EEG subcomponents.

Multivariate backward modelling. In order to assess if and how the location of the moving stimuli was represented in the EEG, we used a multivariate linear reconstruction model g to reconstruct the sound stimulus trajectory S from the neural data R see (Crosse et al., 2016); Bednar and Lalor (2018) for more details. Briefly, this mapping can be described as:

$$\widehat{s}(t) = \sum_{n=1}^{N} \sum_{\tau} g(\tau, n) r(t + \tau, n),$$

where $\widehat{s}(t)$ is the reconstructed estimate of the stimulus position at time $t = 1 \ldots T$, $g(\tau, n)$ is the decoder model which is a function of the time lag τ and the electrode channel $n = 1 \ldots N$, and $r(t, n)$ is the neural response at time t and electrode n. The stimulus trajectory, $\widehat{s}(t)$, is a measure of the azimuth, where an azimuth of $+90°$ corresponds to the right and $-90°$ corresponds to the left (see Fig. 1A). The decoder $g(\tau, n)$ integrates EEG over time lags τ from 0 ms to 250 ms poststimulus to reconstruct each sample of the stimulus trajectory. See Fig. 2. Note that Tikhonov regularization was used to solve ill-posed estimation issue and to prevent model overfitting (equations 9 and 10, Crosse et al., 2016).

Leave-one-out cross-validation was used as a performance measure of our stimulus reconstruction. We fitted the model (decoder) on all but one trial and then evaluated this decoder on the remaining left out trial. Afterwards, we measured the similarity between the reconstructed and the original sound trajectory using Pearson's r. We repeated this for each trial. These correlation coefficients were then averaged across trials for each subject.

To measure the statistical significance of our reconstruction accuracy we used a permutation approach. Specifically, for each decoder, we first established a chance level decoding accuracy by running the leave-one-
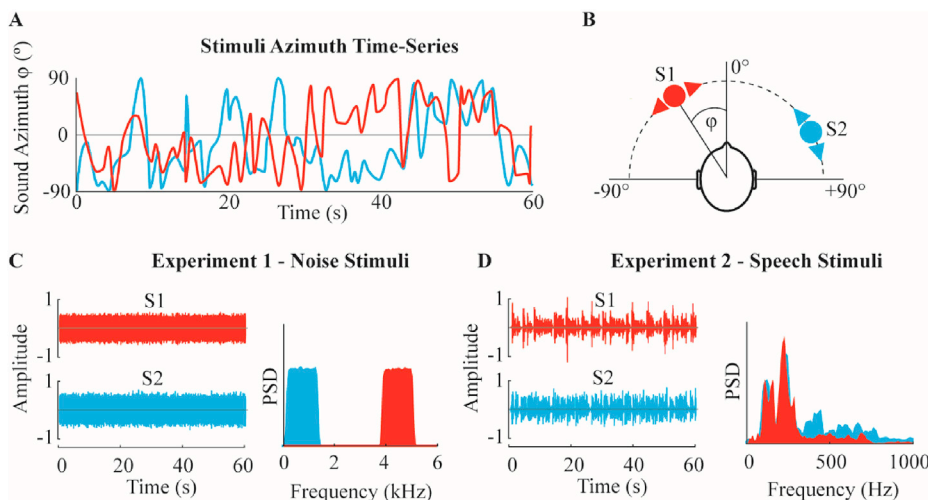

Fig. 1. Continuously moving auditory stimuli. In experiment 1 (noise stimuli) and experiment 2 (speech stimuli), the subjects were presented with two concurrent stimuli, which were HRTF-filtered to be perceived as randomly moving in the frontal part of the horizontal plane between −90 (left) and +90° (right). (A) Example of stimulus trajectories shown as azimuth time-series. (B) Top view of the auditory scene. (C) Left: Example waveforms of band-passed filtered noise stimuli that were used in Experiment 1. Right: Power spectrum density (PSD) shown for Low and High noise stimuli having bandwidth of 1 kHz and center frequencies of 750 Hz and 4500 Hz respectively. (D) Left: Example waveforms of speech stimuli that were used in Experiment 2. Right: PSD shown for both speech waveforms.
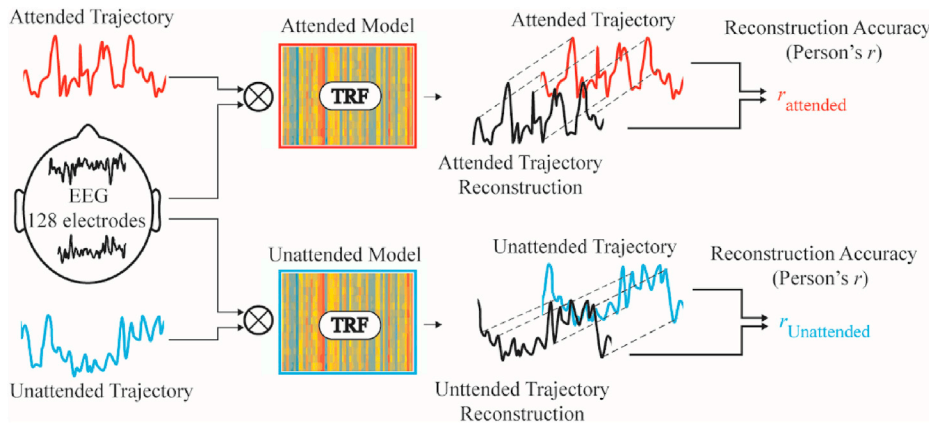
Fig. 2. Decoding EEG analysis. For both attended and unattended sound stimuli, we trained regression models to separately reconstruct the attended and unattended trajectory from EEG. The reconstruction models were using simultaneously information from all EEG channels and time lags corresponding to time window 0–250 ms. The models were trained using leave-one-out cross-validation approach and the reconstruction accuracy of both trajectories was assed using Pearson's r by correlating the actual and reconstructed trajectories.

out cross-validation process 100 times with randomly permuted trial labels. After, for each subject, we computed the chance-level accuracy by averaging the permutation test correlations across all trials and repetitions. Finally, we used one-sided Wilcoxon signed-rank comparison to test whether the "true" reconstruction accuracy is significantly above the computed chance level. All other group-level comparisons between decoders were done using two-sided Wilcoxon signed-rank test.

We also repeated the above analysis where we focused on reconstructing the spatial trajectory of one stimulus, while controlling for (i.e., regressing out) the effect of the other stimulus on the EEG data. To do this we used linear predictive (forward) models (see Crosse et al., 2016): We first trained a forward model to predict EEG from a given stimulus trajectory. Next, we subtracted that predicted EEG trajectory from the original EEG signal. Then we fed the residual EEG data into our stimulus trajectory reconstruction analysis described above.

As the backward model weights are not readily interpretable in a neurophysiological sense, we transformed them into forward "activation patterns", which are more interpretable in terms of the underlying physiology (equation 6, Haufe et al., 2014):

$$A = \Sigma_r g \Sigma_{\hat{s}}^{-1}$$

where A is the activation pattern, Σ_r and $\Sigma_{\hat{s}}$ are covariance matrices corresponding to neural response r and reconstructed stimulus $\hat{s}$, and g are the decoder weights.

Decoding selective attention. In the second experiment, we attempted to use our trajectory reconstruction method to decode selective attention i.e. to determine which of the two speakers was attended from the EEG data. The traditional approach for determining the attended speech stream has been based on stimulus envelope reconstruction. Briefly, the stimulus envelope is reconstructed from the EEG and compared with actual attended and unattended speech envelopes using Pearson's r, then the envelope which has the largest correlation with the reconstructed envelope is deemed the attended envelope (e.g. O'Sullivan et al., 2015).

Here, we used a similar approach as O'Sullivan, with a main difference that we were reconstructing sound trajectory instead of sound envelope: First, we used a leave-one-out cross-validation approach to train an "attended model", which reconstructed the attended stimulus trajectory from the EEG, and used this model to reconstruct the trajectories of the left-out trials. Afterwards, for each subject and each trial, we compared the reconstructed trajectory with the actual attended and unattended stimulus trajectories by calculating Pearson's r, which we will refer to as $r_{\text{traj-attended}}$ and $r_{\text{traj-unattended}}$. We then used a linear discriminant analysis (LDA), which used $r_{\text{traj-attended}}$ and $r_{\text{traj-unattended}}$ as input features, to determine whether the given trial was attended or unattended. Again, this was done using a leave-one-out cross-validation approach. Afterwards, the single-subject decoding accuracy was

calculated as the proportion of correctly classified trials. This was done for Delta and Alpha Pwr models separately. Note that one could decode selective attention just by taking difference between $r_{\text{traj-attended}}$ and $r_{\text{traj-unattended}}$ and consider a trial to be correctly decoded if $r_{\text{traj-attended}} > r_{\text{traj-unattended}}$. Here, to keep the decoding method consistent with combined decoders (see below), we chose to use LDA. This did not impact the decoding accuracy.

In addition, to get a better idea how the performance of trajectory-based attention decoders compares to envelope-based methods, we use the above mentioned approach to train an envelope decoder on the current dataset. That is, we repeated the analysis using the stimulus envelope instead of the stimulus trajectory and ran the LDA using $r_{\text{env-ttended}}$ and $r_{\text{env-unattended}}$, values corresponding to the correlation values between the reconstructed envelope and the actual attended and unattended envelopes, as input features. As done by O'Sullivan et al. (2015), this decoding was performed on EEG that was filtered between 2 and 8 Hz.

Finally, we also tried to combine both the trajectory and envelope based attention decoders. To do so, we ran the LDA decoder using an input feature vector containing 6 correlation values: two pairs of $r_{\text{traj-ttended}}$ and $r_{\text{traj-unattended}}$ corresponding to the Delta and Alpha trajectory decoders and $r_{\text{env-ttended}}$ and $r_{\text{env-unattended}}$ from the envelope decoder.

To statistically assess whether classification performance is above theoretical chance on a single-subject level, we used a nonparametric permutation test (Combrisson and Jerbi, 2015). First, to establish a null distribution of the classification accuracies, we repeated the LDA 1000 times with randomly permuted classification labels. Afterwards, we used the tail of this empirical distribution to calculate the p-value for the original classification. The group level statistical comparison was conducted using one-sided Wilcoxon signed-rank test by testing the actual classification accuracy of each subject against the chance level accuracy that was obtained by averaging the null-distribution.

3. Results

3.1. Experiment 1 – noise stimuli

We investigated whether the spatial dynamics of multiple moving sound sources are separately reflected in scalp recorded EEG and were interested in how attention affects the neural tracking of those spatial dynamics. Specifically, we attempted to decode the trajectories of two simultaneously presented noise stimuli from EEG in a situation when one source was attended and the other unattended. Our hypothesis was that the delta phase and alpha power components of EEG might be differentially sensitive to bottom-up effects induced by spatial properties of the stimuli and top-down attentional selection effects.

Behavioral results. During the experiment, subjects were asked to indicate the spatial location of tremolo targets within the attended audio stream. This was done to confirm that subjects attended the correct

stimuli and to quantify how well subjects perceived the spatial aspects of the auditory scene. Subjects responded quite accurately to this challenging task with 72.8% of attended targets correctly detected. And they were successfully able to ignore the unattended with only 4.7% of targets in that stream being responded to as false alarms. The group-average sensitivity d-prime index was d = 2.08. We also calculated d-prime sensitivity indices separately for low (d = 1.67) and high stimuli (d = 1.51) and found no significant differences.

Trajectory reconstruction results. First, we investigated what EEG signal frequency bandwidth would be the most informative for sound trajectory reconstruction. To test this, we filtered the EEG into different frequency bands and for each band we separately ran the trajectory reconstruction models. This was performed on EEG amplitude signal as well as on EEG power. Consistent with our previous single-source study (Bednar and Lalor, 2018), for both attended and unattended EEG the delta (0.05–2 Hz) and the alpha power (8–12 Hz) of EEG were the most informative for sound trajectory reconstruction (see Fig. 3). In further analysis, we therefore focused on these two EEG subcomponents. Specifically, we individually trained the trajectory reconstruction models on the delta phase and the alpha power of EEG, naming the models "Delta" and "Alpha Pwr" respectively.

Using both Delta and Alpha Pwr decoders, we found that we could reconstruct the sound trajectory of the attended noise source with accuracies above the chance level. For the attended sound source, the average reconstruction correlations were r = 0.079, p = 1.6e-4 and r = 0.055, p = 1.9e-4 for Delta and Alpha Pwr respectively. For the unattended sound source, the reconstruction accuracies were below the chance level, r = 0.01, p = 0.56 for the Delta decoder and r = 0.013, p = 0.13 for the Alpha Pwr decoder. The trajectory reconstruction accuracies showed strong effects of attention. Specifically, the reconstruction correlation values were larger for the attended sound source than the unattended for Delta (p = 3e-4) as well as the Alpha Pwr of the EEG (p = 6e-4). See left panel in Fig. 4A. Also, see examples of reconstructed trajectories in Fig. 4B.

In our experiment, the attended and unattended source trajectories shared a common location at times. In order to minimize any confounding effects that location overlap between the sources might have on our assessment of the reconstructed trajectories, we ran the decoders on the EEG that had the unattended trajectory partialled-out, i.e., that contained only spatial information about the attended source. Similarly, we evaluated the decoders on EEG that had regressed-out the attended trajectory (see methods section for details). When using EEG data with partialled-out unattended source, the trajectory reconstruction accuracy of the unattended source decreased to zero (trivially) and the reconstruction accuracy of the attended source did not significantly change. This pattern was the same for the Delta and the Alpha Pwr decoders. See

middle panel in Fig. 4A.

When we removed the attended trajectory information, the reconstruction accuracy of the attended trajectory decreased below the chance level (p < 0.05). The accuracy of the unattended source trajectory reconstruction increased on average, however it remained below the significance threshold of p < 0.05. Again, this effect was the same in both Delta and Alpha decoders. See right panel in Fig. 4A.

We then ran the decoders at different time-lags between the EEG and the stimulus trajectory using a sliding window of 50 ms. We found that the attended source decoding peaked at around 200 and 125 ms for Delta and Alpha Pwr decoders respectively. For the unattended source, the single-lag accuracy was below the significance level threshold. As discussed in Bednar and Lalor (2018), due to the relatively low velocity of the sound stimuli, the reconstruction accuracy remains high over a relatively large time-lag interval.

For selected time-lags, we plotted the decoder activation patterns (see Fig. 5 below). For the attended source, the Delta decoder showed strong activations over temporal scalp bilaterally. The activation patterns were opposite between the hemispheres, i.e., an increase in activity in one hemisphere was accompanied by a decrease in the other. The Alpha Pwr decoder indicated more posterior activation in parieto-occipital areas. The stimulus caused a relative decrease in activity contralateral to the stimulus position, and a relative increase over ipsilateral scalp. These activation patterns indicating activity over fronto-temporal areas strongly resemble the patterns we observed in our previous study that involved listening to single moving sound source (Bednar and Lalor, 2018). For the unattended decoder, although the activation patterns were also opposite between the hemispheres, the spatial distribution of activation patterns differed from the attended decoders and were generally noisier. The Delta decoder patterns were again more frontal and indicated a relative decrease in activity contralateral to the stimulus position. The Alpha Pwr unattended decoder patterns were localized more posteriorly and were more central in comparison to the attended Alpha Pwr decoder. However, one needs to be careful to interpret these patterns as the reconstruction accuracies of these models were below the chance level. Nevertheless, there was still some structure in the decoder weights, which might suggest some weak cortical tracking of the unattended source.

Correlation with behavioral results. We compared the behavior and trajectory reconstruction data across subjects and found there was no significant correlation between the target detection sensitivity and trajectory reconstruction of the attended or unattended stimuli from the EEG data (all p > 0.05).

3.2. Experiment 2- speech stimuli

In the second experiment, we were interested in whether we could employ a more naturalistic attention paradigm and attempted to reconstruct the trajectories of two concurrent non-stationary speech sources (speakers). Also, in contrast to the first experiment, the subjects performed a non-spatial speech comprehension task. This was done to see if our spatial decoding measures were still affected by attention even when the task wasn't specifically spatial. Finally, we tried using trajectory reconstruction to decode selective attention in multi-speaker environment and compare it with 'traditional' envelope-based cocktail party decoders. In addition, we tried to see if markers of spatial attention could be added to traditional cocktail party measures to improve the selective attention decoding performance.

Behavioral results. In the second experiment, subjects were asked to answer multiple-choice questions on the attended and unattended story after each trial. On average, the subjects answer correctly 69.5% (p < 0.005) of the questions from the attended story and 25.1% from the unattended story, which was not greater than chance (25%, p > 0.05). The number of correctly answered question was significantly higher for the attended than the unattended stream (p < 0.005). Note that the study by (O'Sullivan et al., 2015), which employed dichotic stimulation of the
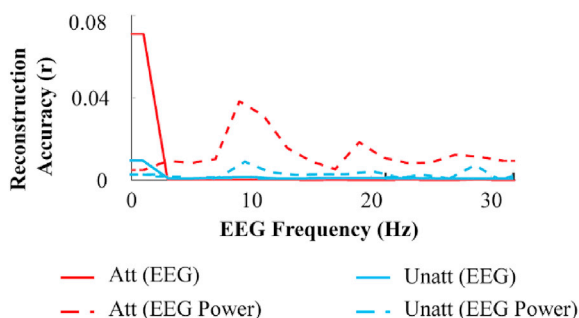

Fig. 3. Experiment 1- Noise Stimuli: Trajectory decoding performance dependency the EEG bandwidth. The trajectory reconstruction accuracies shown for different frequencies of raw EEG signal (solid) and EEG power (dashed) for the attended (red) or unattended (blue) stimuli. The EEG power was calculated as an absolute value of a Hilbert transform of a raw EEG signal. Different frequencies of the signals were obtained using a sliding band-pass filter with pass-band width of 2 Hz.
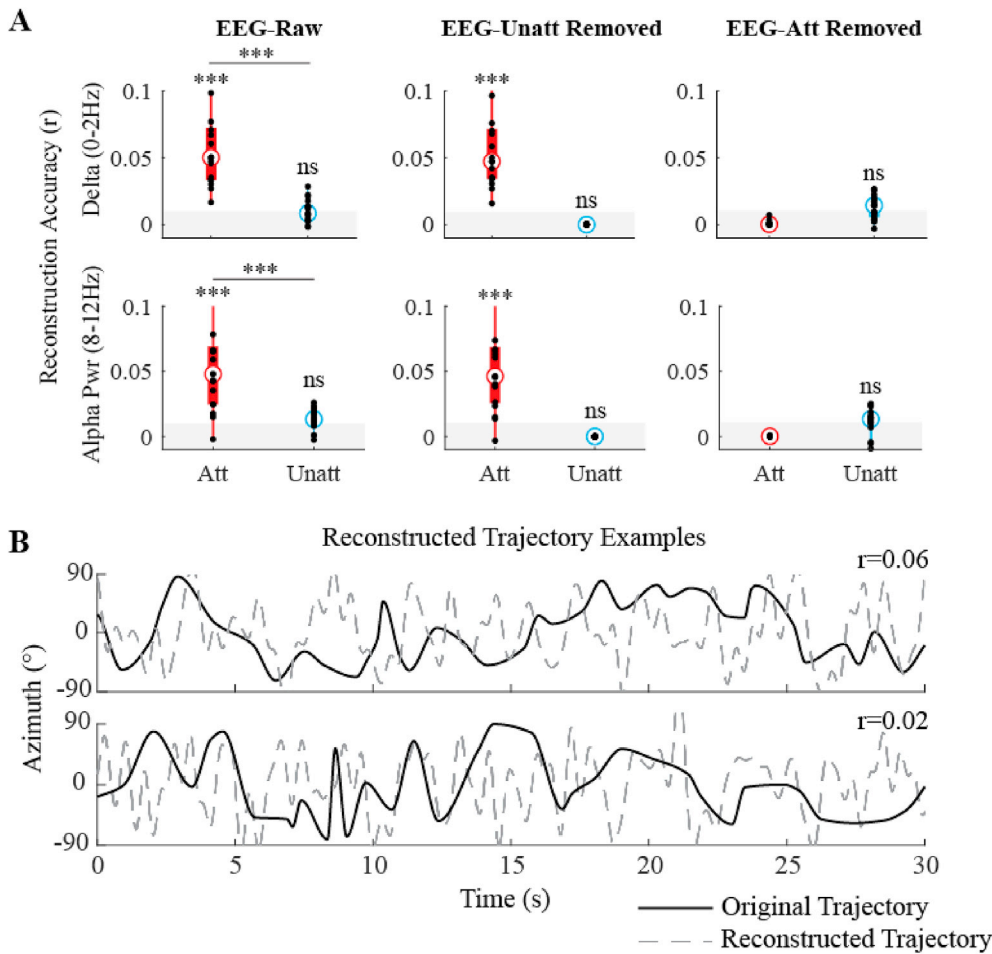
Fig. 4. Experiment 1- Noise Stimuli: Reconstructing sound trajectories of two continuously moving noise sources from EEG. Stimulus trajectory reconstruction accuracies are shown for the attended ("Att", red) and unattended stimuli ("Unatt", blue) for Delta and Alpha Pwr decoders. The reconstruction was performed on: (1) filtered-only EEG ("Raw"), (2) filtered EEG with partialled-out unattended stimulus trajectory ("Unatt Removed") and (3) filtered EEG signal with partialled-out attended stimulus trajectory ("Att Removed"). Shaded gray area at the bottom of each plot indicates significance threshold ($p < 0.05$). * ** *** indicates reconstruction accuracies at the level of $p < 0.05$, $p < 0.01$ and $p < 0.005$ respectively. (B) Examples of original sound stimulus trajectories along with their reconstructions shown for a trial segments of one subject. The correlation values between the original and reconstructed trajectories are shown above the plots.

same speech stimuli and the identical set of questions, reported comprehension accuracy of 80.4%. This is approximately 10% higher accuracy than we found here. We believe that this might be due to increased difficulty of the task caused by the continuous movement of the stimuli.

Trajectory reconstruction results. Using the Delta decoder, we found we could successfully reconstruct the trajectory of the attended and unattended speech sources with reconstruction correlation values of $r = 0.057$, $p = 4.8e-3$ and $r = 0.024$, $p = 0.032$ respectively. For the Alpha Pwr decoder, the correlation values for the attended and unattended stimuli were $r = 0.042$, $p = 4.2e-3$ and $r = 0.019$, $p = 0.5$. The reconstruction values were significantly larger for the attended than the unattended speech stimuli only for the Delta EEG $p = 0.019$ but not for the Alpha Pwr decoder $p = 0.065$. See Fig. 6 for reconstruction results.

As in the first experiment, we attempted to reconstruct the attended and unattended sources in isolation by regressing-out the other source. Again, the reconstruction accuracies of the attended sound source were unaffected when we removed the trajectory information of the unattended source (see Fig. 6, middle panel). Importantly, when we removed the positional information of the attended sound source from the EEG, it was still possible to reconstruct the unattended source trajectory using the Delta decoder ($r = 0.023$, $p = 0.032$). This indicates that the above chance-level reconstruction accuracy of the unattended sound source trajectory was not simply caused by the time intervals where both sound sources were co-located.

Fig. 7 shows the activation patterns of the speech trajectory decoders and the single-lag decoding results. The attended decoder patterns indicate strong contralateral tuning and opposite activations between hemispheres and their morphology resembles the activation patterns from the first experiment.

The activation patterns corresponding to the unattended decoder were relatively noisy and interestingly less lateralized than the attended decoder. Again, this should be interpreted with caution since the reconstruction accuracy of the model was weak. The single-lag decoding showed that the trajectory of the attended sound is best decoded at 195 and 250 ms for Delta and Alpha decoders respectively. In comparison to the first experiment, the Alpha decoding peaked around 125 ms later, which might be possibly attributed to the different nature if the attention task.

Decoding selective attention at single-trial level using sound trajectory and envelope reconstruction. As the trajectory reconstruction appeared to be strongly modulated by selective attention, we tested the possibility of using the spatial decoders to decode selective attention in a cocktail party, i.e., to decode which of the two speakers was attended. We used an approach similar to (O'Sullivan et al., 2015), who used a method based on sound amplitude (envelope) reconstruction to decode selective attention from EEG at a single trial level.

Briefly, our method was to train a trajectory reconstruction model on the attended speech stimulus. Next, using a previously unseen EEG trial, we reconstructed the stimulus trajectory and assessed the correlation between the predicted trajectory and each of the two original speech stimulus trajectories. Finally, we used LDA to classify the correlation values as attended or unattended speech stream. As well as this trajectory-based decoding, we also used the same dataset to perform envelope-based decoding of selective attention, which is using the same regression approach, and compared the results. Finally, we also tried to combine the trajectory- and envelope-based decoder together. See methods section for details.

On a group-level, the trajectory-based approach decoded selective attention with an accuracy significantly above the chance level for the
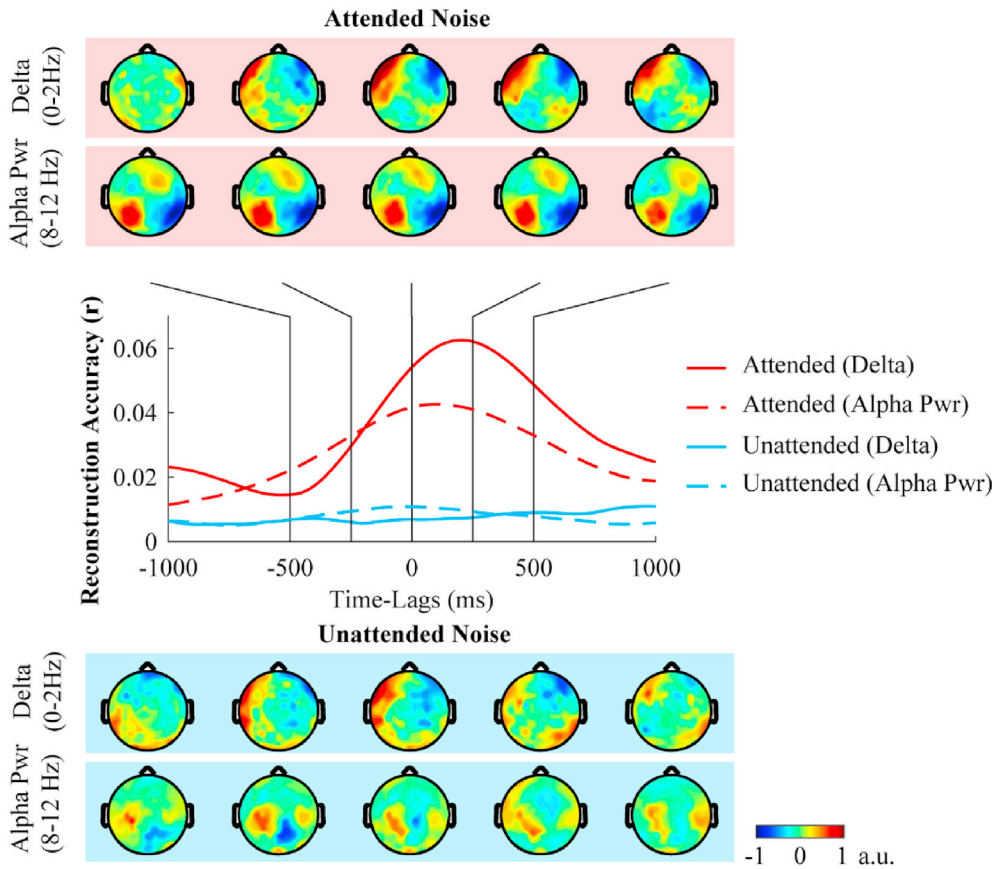
Figure 5. Experiment 5- Noise Stimuli: Trajectory decoder activation patterns and single-lag decoding. The decoders were independently evaluated at single time-lags between the EEG and the trajectory signal. Middle: Reconstruction accuracies shown for each individual time-lag. Top and Bottom: Decoder activation patterns corresponding to different time-lags are shown for the attended (red background) and unattended speech stimuli (blue background). The decoder patterns for each time-lag were individually normalized.
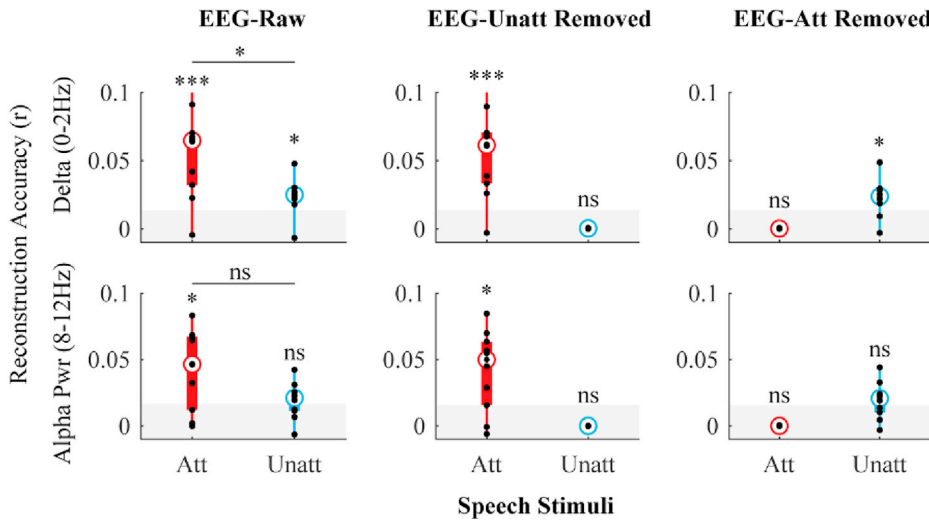

Fig. 6. Experiment 2- Speech Stimuli: Reconstructing sound trajectories of two moving speech sources from EEG. (A) Stimulus trajectory reconstruction accuracies shown for the attended ("Att") and unattended stimuli ("Unatt") for Delta and Alpha Pwr decoders. The reconstruction was performed on: (1) filtered-only EEG ("Raw"), (2) filtered EEG with partialled-out unattended stimulus trajectory ("Unatt Removed") and (3) filtered EEG signal with partialled-out attended stimulus trajectory ("Att Removed"). Shaded gray area at the bottom of each plot indicates significance threshold (p < 0.05). * ** *** indicates reconstruction accuracies at the level of p < 0.05, p < 0.01 and p < 0.005 respectively.

Delta decoder with an accuracy of 66.75% (p = 2e-3), but not for the Alpha Pwr decoder, which reached 53.75% (p = 0.28). The average decoding accuracy of the envelope decoder was 92% (p = 9.8e-4) and was better than both trajectory decoders (p < 0.005). Finally, the decoder that combined both trajectory decoders and envelope decoder reached average decoding accuracy of 93% (p = 9.8e-4). This was on average higher decoding accuracy than for envelope decoder alone, however, the difference was not significant (p = 0.53). Within-subject statistical analysis showed that for the Delta and Alpha Pwr trajectory decoders, the decoding accuracy was above the chance level for 6/10 and 3/10 subjects respectively. For the envelope and combined decoder, the single subject data showed significant decoding accuracy for 10/10 subjects. See results

in Fig. 8.

4. Discussion

Here in this study, we were interested in whether we can decode sound source locations from EEG when listeners were presented with two concurrent moving stimuli and attended to one of them. Specifically, we tried to reconstruct the trajectory of attended as well as unattended stimuli. This was tested in two experiments which differed in their stimulus type and behavioral task: Experiment 1 employed filtered noise stimuli that were discriminable by their frequency content and subjects performed a spatial task. Experiment 2 involved listening to two different
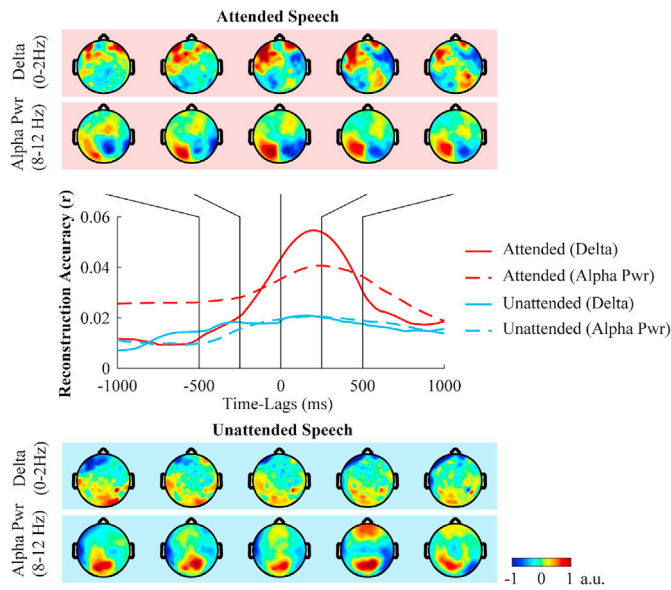
Fig. 7. Experiment 2- Speech Stimuli: Trajectory decoder activation patterns and single-lag decoding. The decoders were independently evaluated at single time-lags between the EEG and the trajectory signal. Middle: Reconstruction accuracies shown for each individual time-lag. Top and Bottom: Decoder activation patterns corresponding to different time-lags are shown for the attended (red background) and unattended speech stimuli (blue background). The decoder patterns for each time-lag were individually normalized.

speakers telling a story and a speech comprehension task.

The trajectory of an attended sound can be reconstructed from EEG even in the presence of another stimulus. Previously, it has been shown that a single-source location can been decoded from recorded neurophysiological data (Zhang et al., 2015; Derey et al., 2016; Bednar et al., 2017). Moreover, in our previous study (Bednar and Lalor, 2018), we showed that in an acoustic environment with one continuously moving sound source, the cortex dynamically tracks the source's location in time and demonstrated that the trajectory of the source can be reconstructed using a regression model. Importantly, we also demonstrated that the trajectory can be successfully applied to the ITD-spatialized pulse-train, which shows that our trajectory reconstruction is not purely driven by envelope fluctuations.

Here in this paper, we showed that in a more complex acoustic environment with two continuously moving sources, the sound trajectory of the attended sound source can be successfully decoded from EEG. As in our previous study, we identified two distinct EEG components that tracked the location of the attended sound: (1) low frequency delta (0–2 Hz) component, which showed strong bilateral activation indicating

involvement of auditory cortex and (2) the alpha power (8–12 Hz) of EEG that was over the parieto-occipital region. The cortical activation within the delta band indicated the hemispheric preference for contralateral stimuli as shown previously (e.g. Palomaki et al., 2000; Palomaki et al., 2005). The alpha EEG activation pattern looked similar to patterns observed in studies that involved deploying spatial attention to auditory stimuli (Kerlin et al., 2010; Wöstmann et al., 2016).

Active sound localization task is not required to successfully reconstruct the attended sound trajectory. It has been shown that paying attention to the location of a sound enhances neural tuning to this stimulus feature (Zatorre et al., 2002; Ahveninen et al., 2006b; Altmann et al., 2008). In the first experiment, we successfully decoded the sound trajectory when subjects performed a spatial task. However, it was unclear whether it is possible to apply our decoding framework in situations where one does not perform an active sound localization task, as in this case one might expect reduced spatial sensitivity of the cortex. However, as we showed in our second experiment, which involved a non-spatial speech comprehension task, it is possible to reconstruct the attended sound trajectory even when subjects were not tasked with localizing targets. This is in line with observations from an fMRI study by Deouell et al. (2007) that demonstrated that the cortex is spatially sensitive even when subjects do not perform a localization task.

Do we track the position of the unattended sound? Previous studies showed that we are sensitive to spatial separation between sound sources within the auditory scene (Zatorre et al., 2002; Shiell et al., 2018). However, it has been unclear whether the exact location of the unattended source is represented within the auditory cortex. We tested this by attempting to reconstruct the unattended sound source trajectory from the EEG signal.

The data from our first experiment, which involved noise stimuli, showed that the unattended sound source position cannot be successfully reconstructed from the EEG signal. However, in the second experiment that involved speech, the results indicated that the unattended source position can be successfully decoded using the delta EEG component. As one might expect, the unattended reconstruction accuracies were lower than for the attended stimulus, which is in line with studies that showed that the selective attention enhances the neural representation of the attended sound (Woldorff et al., 1993; Petkov et al., 2004).

However, in both of our experiments, the trajectories of attended and unattended sources were partially correlated and at times shared the same location. Therefore, we were concerned that the decoding of the unattended sound source was possibly driven by the moments when the locations of both sound sources coincided. In order to test this possibility, we first regressed-out the attended trajectory from the EEG signal and then attempted to reconstruct the unattended sound source location. The results showed that it is still possible to reconstruct the unattended speech stimulus trajectory with a decoding accuracy above the chance level, which further supports the idea that cortex actually tracks the
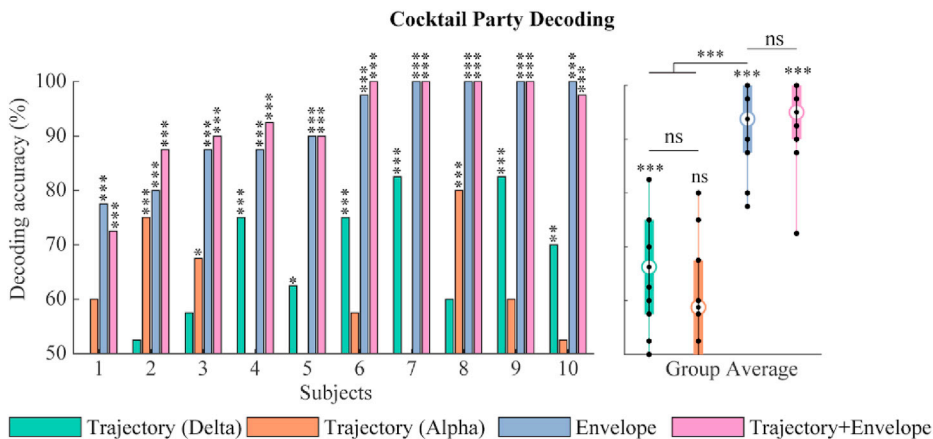

Fig. 8. Experiment 2- Speech Stimuli: Decoding selective attention using trajectory and envelope reconstruction models. The decoding was performed using Delta and Alpha Pwr trajectory decoders, envelope decoder and "Trajectory + Envelope" decoder that integrated both trajectory decoders and envelope decoder together. The decoding accuracy is shown for individual subjects (left) as well as group average (right). Subjects are sorted according to the performance of their envelope decoder. * ** *** indicates prediction differences at the level of p < 0.05, p < 0.01 and p < 0.005 respectively.

unattended speech source trajectory.

The changes in spatial sound position and related changes in interaural level differences caused smooth amplitude envelope fluctuations of the sound signal. Therefore, it is possible that our trajectory reconstruction was driven by the fluctuating envelope. Nevertheless, we think that the successful reconstruction of the unattended sound source trajectory represents spatial position tracking and not the sound envelope tracking. This assumption is based on our previous study (Bednar and Lalor, 2018), which demonstrated that we can reconstruct the position of a single moving sound source, which was a pulse train that contained only interaural time difference (ITD) cues and did not have a fluctuating amplitude envelope that changed with source position.

Location tracking by delta and alpha power EEG possibly represent different mechanisms. We predicted that the Delta decoder might track attended as well as unattended trajectories and we expected that the unattended tracking would be weaker. This assumed that the delta EEG trajectory tracking would directly encode the spatial location of the stimulus and that would just be modulated by attention, similar to amplitude envelope tracking by sub-alpha band (<8hz) EEG, which can be seen in cocktail party studies (e.g. O'Sullivan et al., 2015) that showed this pattern. In contrast, for the alpha power EEG component, which is often associated with spatial attention per se (Kerlin et al., 2010; Wöstmann et al., 2016), we hypothesized that it would track only the attended sound source trajectory. Indeed, the results from our second experiment, which employed speech stimuli, support this dissociation of delta and alpha power components. However, for the noise stimuli (experiment 1), the decoding accuracy of the unattended sound source was below the detection level for both frequency components of EEG.

A question remains then as to what the delta and alpha power EEG components actually represent in the context of sound location tracking. The morphology of the delta EEG reconstruction model indicated enhanced cortical responses contralaterally to the stimulus location. This suggests that the delta EEG response reflects intrinsic contralateral tuning of the auditory system that has been described in other studies (e.g. Palomaki et al., 2000; Lewald and Getzmann, 2011; Briley et al., 2013; Derey et al., 2016). With respect to the alpha EEG component, as we mention above, the lateralized posterior scalp alpha power is often related to auditory spatial attention. In line with this, we found that the alpha tracking is restricted to the attended sound source. This might suggest that our alpha power-based reconstruction reflects allocation of spatial attention rather than the spatial location of the sound source per se. Indeed the parietal distribution of our alpha power decoder weights supports the notion that this might reflect auditory spatial attention (Banerjee et al., 2011). However, we cannot completely rule out the possibility of a cross-modal effect with our auditory stimulus eliciting covert visual attention effects. In line with this idea, a recent study by Feng et al. (2017) showed that a spatial auditory cue caused posterior alpha lateralization and further demonstrated that the changes in alpha activity predicted the accuracy in subsequent visual task. The authors also speculated that the posterior alpha desynchronization might reflect a general priming mechanism that facilitates visual processing.

Is the trajectory tracking motion specific? It has been discussed whether the moving sounds engage different cortical networks than static sounds. Neuroimaging studies have indicated that non-primary auditory cortex and particularly planum temporale (PT) is involved motion processing (Baumgart et al., 1999; Pavani et al., 2002; Warren et al., 2002; Krumbholz et al., 2005a). However, these structures often respond to spatial static sounds and so it is debatable as to whether they represent true motion detectors. Nevertheless, a recent study provided strong evidence for cortical motion sensitivity and showed that activity in posterior belt and parabelt cortical regions cannot be explained by static spatial and spectro-temporal processes (Poirier et al., 2017). Although the cortical motion sensitivity might have contributed to our trajectory reconstruction, we speculate that the spatial tracking as we observed is not motion specific. As we discussed above, the contralateral tuning of the cortex (Palomaki et al., 2000, 2005; Magezi and Krumbholz, 2010;

Briley et al., 2013) as well as alpha lateralization (Kerlin et al., 2010; Wöstmann et al., 2016; Feng et al., 2017) have been mainly demonstrated for static sound stimuli.

Different spatial sensitivity to moving noise and speech stimuli. It has been proposed, that in a multi-source environment, listeners are able to collect spatial information from both sources during "glimpses" where sound envelopes do not overlap and this allows listeners to track the position of multiple sources (Yost and Brown, 2013). This could explain why we were unable to reconstruct the unattended sound trajectory in the first experiment, which used noise stimuli, and why we were successful in the second experiment that employed speech. Specifically, in the first experiment, the noise stimuli had relatively constant intensity and were ongoing during the whole trial length, which possibly did not allow to listeners to "collect" enough information about the unattended sound source. In contrast, in the second experiment that employed competing speech, the stimuli had varying intensity included periods of silence, e.g., breaks between words and sentences. Therefore, subjects could possibly utilize these glimpses to tracks both sources. Another possibility as to why we were able to decode the location of the unattended speech but not the unattended noise stimulus is that it is easier to localize and follow speech than filtered noise due to having access to more spatial acoustic cues, as it has been previously shown that it is easier to localize amplitude modulated signals (Blauert, 1997).

Potential for using stimulus trajectory reconstruction to decode selective attention. With more practical applications in mind, we tested whether one can use the trajectory reconstruction approach to build a decoder of selective attention, which can be used in cognitively steered hearing aids (Lunner et al., 2009; Mirkovic et al., 2015). We used the same approach as the envelope-based reconstruction decoder, which is capable of successfully solving the "cocktail party problem" on a single-trial level (O'Sullivan et al., 2015). We demonstrated that it is possible to use the trajectory reconstruction approach to decode to which speaker the listener was attending with above-chance accuracy. However, note that this method assumes that the location of speakers are known, which might not be easily achievable in practice.

We also compared the results with the envelope-based decoder of selective attention. By a clear margin, the envelope-based model outperformed the trajectory reconstruction model, showing that the neural tracking of the attended stimulus envelope is more robust than that of the stimulus location. We also tried to combine the trajectory and envelope decoders together, however, although the combined decoder led to slightly better average accuracy, this gain was not significant. Nevertheless, unlike the envelope decoder, the trajectory reconstruction model directly predicts the sound source angle, which can be theoretically used as control signal for cognitively steering hearing aids. Therefore it might be of interest in future studies to investigate the possibility of integrating both types of decoders.

5. Conclusion

We showed the trajectory of an attended sound source can be reliably reconstructed from EEG data and that the reconstruction results are robust to the presence of distracting auditory stimuli and the specific behavioral task. With respect to reconstructing the trajectory of the unattended sound, our data suggested that the cortical representation of the unattended source position is below detection level for the noise stimuli but we observed weak tracking of the unattended source location for the speech stimuli. Our data indicated that the ability to track the unattended sound source position in time was likely driven by temporal glimpses in which the sound intensities of the attended and unattended sources did not overlap. We also demonstrated that the trajectory reconstruction method can be in principle used to decode selective attention on a single-trial basis, however, its performance is inferior to envelope-based decoders.

Declaration of competing interest

None.

Acknowledgment

References

Ahveninen, J., Jääskeläinen, I.P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., Levänen, S., Lin, F.-H., Sams, M., Shinn-Cunningham, B.G., Witzel, T., Belliveau, J.W., 2006a. Task-modulated "What" and "Where" Pathways in Human Auditory Cortex, 103, pp. 14608–14613.

Ahveninen, J., Jaaskelainen, I.P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F.H., Sams, M., Shinn-Cunningham, B.G., Witzel, T., Belliveau, J.W., 2006b. Task-modulated "what" and "where" pathways in human auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 103, 14608–14613.

Altmann, C.F., Henning, M., Doring, M.K., Kaiser, J., 2008. Effects of feature-selective attention on auditory pattern and location processing. Neuroimage 41, 69–79.

Arnott, S.R., Binns, M.A., Grady, C.L., Alain, C., 2004. Assessing the auditory dual-pathway model in humans. Neuroimage 22, 401–408.

Banerjee, S., Snyder, A.C., Molholm, S., Foxe, J.J., 2011. Oscillatory alpha-band mechanisms and the deployment of spatial attention to anticipated auditory and visual target locations: supramodal or sensory-specific control mechanisms? J. Neurosci. 31, 9923–9932.

Baumgart, F., Gaschler-Markefski, B., Woldorff, M.G., Heinze, H.-J., Scheich, H., 1999. A movement-sensitive area in auditory cortex. Nature 400, 724–726.

Bednar, A., Boland, F.M., Lalor, E.C., 2017. Different spatio-temporal electroencephalography features drive the successful decoding of binaural and monaural cues for sound localization. Eur. J. Neurosci. 45, 679–689.

Bednar, A., Lalor, E.C., 2018. Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG. Neuroimage 181, 683–691.

Best, V., van Schaik, A., Carlile, S.J.T.J.o.t.A.S.o.A., 2004. Separation of Concurrent Broadband Sound Sources by Human Listeners, 115, pp. 324–336.

Blauert, J., 1997. Spatial Hearing: the Psychophysics of Human Sound Localization. MIT press.

Briley, P.M., Kitterick, P.T., Summerfield, A.Q., 2013. Evidence for opponent process analysis of sound source location in humans. J Assoc Res Otolaryngol 14, 83–101.

Brunetti, M., Belardinelli, P., Caulo, M., Del Gratta, C., Della Penna, S., Ferretti, A., Lucci, G., Moretti, A., Pizzella, V., Tartaro, A., Torquati, K., Olivetti Belardinelli, M., Romani, G.L., 2005. Human brain activation during passive listening to sounds from different locations: an fMRI and MEG study. Hum. Brain Mapp. 26, 251–261.

Carl, D., Gutschalk, A.J.E.b.r., 2013. Role of Pattern, Regularity, and Silent Intervals in Auditory Stream Segregation Based on Inter-aural Time Differences, 224, pp. 557–570.

Cherry, E.C., 1953. Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975–979.

Combrisson, E., Jerbi, K., 2015. Exceeding chance level by chance: the caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J. Neurosci. Methods 250, 126–136.

Crosse, M.J., Liberto, D., M, G., Bednar, A., Lalor, E.C., 2016. The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci. 10.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21.

Deouell, L.Y., Heller, A.S., Malach, R., D'Esposito, M., Knight, R.T., 2007. Cerebral responses to change in spatial location of unattended sounds. Neuron 55, 985–996.

Derey, K., Valente, G., de Gelder, B., Formisano, E., 2016. Opponent coding of sound location (azimuth) in planum temporale is robust to sound-level variations. Cerebr. Cortex 26, 450–464.

Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. 109, 11854–11859.

Feng, W., Störmer, V.S., Martinez, A., McDonald, J.J., Hillyard, S.A., 2017. Involuntary orienting of attention to a sound desynchronizes the occipital alpha rhythm and improves visual perception. Neuroimage 150, 318–328.

Furukawa, S., Xu, L., Middlebrooks, J.C., 2000. Coding of sound-source location by ensembles of cortical neurons. J. Neurosci. 20, 1216–1228.

Grothe, B., Pecka, M., McAlpine, D., 2010. Mechanisms of sound localization in mammals. Physiol. Rev. 90, 983–1012.

Harrington, I.A., Stecker, G.C., Macpherson, E.A., Middlebrooks, J.C.J.H.r., 2008. Spatial Sensitivity of Neurons in the Anterior, Posterior, and Primary Fields of Cat Auditory Cortex, 240, pp. 22–41.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110.

Imig, T.J., Irons, W.A., Samson, F.R., 1990. Single-unit selectivity to azimuthal direction and sound pressure level of noise bursts in cat high-frequency primary auditory cortex. J. Neurophysiol. 63, 1448–1466.

Kerlin, J.R., Shahin, A.J., Miller, L.M., 2010. Attentional gain control of ongoing cortical speech representations in a "cocktail party". J. Neurosci. 30, 620–628.

Krumbholz, K., Schönwiesner, M., Rübsamen, R., Zilles, K., Fink, G.R., von Cramon, D.Y., 2005a. Hierarchical processing of sound location and motion in the human brainstem and planum temporale. Eur. J. Neurosci. 21, 230–238.

Krumbholz, K., Schonwiesner, M., von Cramon, D.Y., Rubsamen, R., Shah, N.J., Zilles, K., Fink, G.R., 2005b. Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. Cerebr. Cortex 15, 317–324.

Lee, C.C., Middlebrooks, J.C., 2011. Auditory cortex spatial sensitivity sharpens during task performance. Nat. Neurosci. 14, 108–114.

Lewald, J., Getzmann, S., 2011. When and where of auditory spatial processing in cortex: a novel approach using electrotomography. PLoS One 6, e25146.

Lunner, T., Rudner, M., Rönnberg, J., 2009. Cognition and hearing aids. Scand. J. Psychol. 50, 395–403.

Magezi, D.A., Krumbholz, K., 2010. Evidence for opponent-channel coding of interaural time differences in human auditory cortex. J. Neurophysiol. 104, 1997–2007.

McLaughlin, S.A., Higgins, N.C., Stecker, G.C., 2015. Tuning to binaural cues in human auditory cortex. J Assoc Res Otolaryngol 1–17.

Mesgarani, N., Chang, E.F., 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485, 233.

Middlebrooks, J.C., Bremen, P., 2013. Spatial stream segregation by auditory cortical neurons. J. Neurosci. 33, 10986–11001.

Middlebrooks, J.C., Onsan, Z.A., 2012. Stream segregation with high spatial acuity. J. Acoust. Soc. Am. 132, 3896–3911.

Middlebrooks, J.C., Pettigrew, J.D., 1981. Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. J. Neurosci. 1, 107–120.

Mirkovic, B., Debener, S., Jaeger, M., De Vos, M., 2015. Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. J. Neural Eng. 12, 046007.

Mognon, A., Jovicich, J., Bruzzone, L., Buiatti, M., 2011. ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. Psychophysiology 48, 229–240.

O'Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A., Lalor, E.C., 2015. Attentional selection in a cocktail party environment can Be decoded from single-trial EEG. Cerebr. Cortex 25, 1697–1706.

Ortiz-Rios, M., Azevedo, F.A., Kusmierek, P., Balla, D.Z., Munk, M.H., Keliris, G.A., Logothetis, N.K., Rauschecker, J.P., 2017. Widespread and opponent fMRI signals represent sound location in macaque auditory cortex. Neuron 93, 971–983 e974.

Palomaki, K., Alku, P., Makinen, V., May, P., Tiitinen, H., 2000. Sound localization in the human brain: neuromagnetic observations. Neuroreport 11, 1535–1538.

Palomaki, K.J., Tiitinen, H., Makinen, V., May, P.J., Alku, P., 2005. Spatial processing in human auditory cortex: the effects of 3D, ITD, and ILD stimulation techniques. Brain Res Cogn Brain Res 24, 364–379.

Pavani, F., Macaluso, E., Warren, J.D., Driver, J., Griffiths, T.D., 2002. A common cortical substrate activated by horizontal and vertical sound movement in the human brain. Curr. Biol. 12, 1584–1590.

Petkov, C.I., Kang, X., Alho, K., Bertrand, O., Yund, E.W., Woods, D.L., 2004. Attentional modulation of human auditory cortex. Nat. Neurosci. 7, 658.

Poirier, C., Baumann, S., Dheerendra, P., Joly, O., Hunter, D., Balezeau, F., Sun, L., Rees, A., Petkov, C.I., Thiele, A., Griffiths, T.D., 2017. Auditory motion-specific mechanisms in the primate brain. PLoS Biol. 15, e2001379.

Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. 97, 11800–11806.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., Rauschecker, J.P., 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. 2, 1131–1136.

Shiell, M.M., Hausfeld, L., Formisano, E., 2018. Activity in human auditory cortex represents spatial separation between concurrent sounds. J. Neurosci. 38, 4977–4984.

Shinn-Cunningham, B.G., Schickler, J., Kopčo, N., Litovsky, R., 2001. Spatial unmasking of nearby speech sources in a simulated anechoic environment. J. Acoust. Soc. Am. 110, 1118–1129.

Stecker, G.C., Harrington, I.A., Middlebrooks, J.C., 2005. Location coding by opponent neural populations in the auditory cortex. PLoS Biol. 3, e78.

Stecker, G.C., Mickey, B.J., Macpherson, E.A., Middlebrooks, J.C., 2003. Spatial sensitivity in field PAF of cat auditory cortex. J. Neurophysiol. 89, 2889–2903.

Warren, J.D., Zielinski, B.A., Green, G.G.R., Rauschecker, J.P., Griffiths, T.D., 2002. Perception of sound-source motion by the human brain. Neuron 34, 139–148.

Werner-Reiss, U., Groh, J.M., 2008. A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. J. Neurosci. 28, 3747–3758.

Woldorff, M.G., Gallen, C.C., Hampson, S.A., Hillyard, S.A., Pantev, C., Sobel, D., Bloom, F.E., 1993. Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc. Natl. Acad. Sci. 90, 8722–8726.

Wöstmann, M., Herrmann, B., Maess, B., Obleser, J., 2016. Spatiotemporal dynamics of auditory attention synchronize with speech. Proc. Natl. Acad. Sci. U.S.A. 113, 3873–3878.

Yost, W.A., Brown, C.A., 2013. Localizing the sources of two independent noises: Role of time varying amplitude differences. J. Acoust. Soc. Am. 133 (4), 2301–2313.

Zatorre, R.J., Bouffard, M., Ahad, P., Belin, P., 2002. Where is' where'in the human auditory cortex? Nat. Neurosci. 5, 905–909.

Zhang, X., Zhang, Q., Hu, X., Zhang, B., 2015. Neural representation of three-dimensional acoustic space in the human temporal lobe. Front. Hum. Neurosci. 9, 203.