Behavioral/Cognitive

Decoding Object-Based Auditory Attention from Source-Reconstructed MEG Alpha Oscillations

Ingmar E. J. de Vries, Giorgio Marinato, and Daniel Baldauf
Centre for Mind/Brain Sciences (CIMeC), University of Trento, 38068 Rovereto, Italy

How do we attend to relevant auditory information in complex naturalistic scenes? Much research has focused on detecting which information is attended, without regarding underlying top-down control mechanisms. Studies investigating attentional control generally manipulate and cue specific features in simple stimuli. However, in naturalistic scenes it is impossible to dissociate relevant from irrelevant information based on low-level features. Instead, the brain has to parse and select auditory objects of interest. The neural underpinnings of object-based auditory attention remain not well understood. Here we recorded MEG while 15 healthy human subjects (9 female) prepared for the repetition of an auditory object presented in one of two overlapping naturalistic auditory streams. The stream containing the repetition was prospectively cued with 70% validity. Crucially, this task could not be solved by attending low-level features, but only by processing the objects fully. We trained a linear classifier on the cortical distribution of source-reconstructed oscillatory activity to distinguish which auditory stream was attended. We could successfully classify the attended stream from alpha (8–14 Hz) activity in anticipation of repetition onset. Importantly, attention could only be classified from trials in which subjects subsequently detected the repetition, but not from miss trials. Behavioral relevance was further supported by a correlation between classification accuracy and detection performance. Decodability was not sustained throughout stimulus presentation, but peaked shortly before repetition onset, suggesting that attention acted transiently according to temporal expectations. We thus demonstrate anticipatory alpha oscillations to underlie top-down control of object-based auditory attention in complex naturalistic scenes.

Key words: **auditory attention; MEG source reconstruction; MVPA; naturalistic sound scenes; object-based attention**

Significance Statement

In everyday life, we often find ourselves bombarded with auditory information, from which we need to select what is relevant to our current goals. Previous research has highlighted how we attend to specific highly controlled aspects of the auditory input. Although invaluable, it is still unclear how this relates to attentional control in naturalistic auditory scenes. Here we used the high precision of magnetoencephalography in space and time to investigate the brain mechanisms underlying top-down control of object-based attention in ecologically valid sound scenes. We show that rhythmic activity in auditory association cortex at a frequency of ~10 Hz (alpha waves) controls attention to currently relevant segments within the auditory scene and predicts whether these segments are subsequently detected.

Introduction

How do we select relevant auditory information when faced with distraction in a noisy environment? This question has been commonly referred to as the "cocktail party problem" (Cherry, 1953) and pertains not only to how we attend to one person's speech among others (e.g., at a cocktail party), but, more generally, to selective auditory attention in ecological environments (Shinn-Cunningham, 2008; Ding and Simon, 2012). Ample research has demonstrated the large effect selective attention has on sensory processing of auditory input. Generally, delta-to-theta (i.e., 2–8 Hz) oscillations in auditory cortex track slow acoustic fluctuations (i.e., the temporal envelope) of speech (i.e., entrainment or phase-locking; Giraud and Poeppel, 2012; Ding and Simon, 2014). Importantly, this tracking improves for attended speech (Zion Golumbic et al., 2013; Haegens and Zion Golumbic, 2018), consequently aligning the high-excitability phase of neural oscillations to relevant events in the attended auditory input (Lakatos et al., 2013). While improved tracking unambiguously demonstrates that attention enhances auditory processing, it likely reflects a consequence of attentional selection, rather than top-down control. The neural mechanisms controlling which cortical representation is selected for enhanced processing are far from understood.

Received Mar. 19, 2021; revised Aug. 8, 2021; accepted Aug. 11, 2021.

Author contributions: G.M. and D.B. designed research; I.E.J.d.V., G.M., and D.B. performed research; I.E.J.d.V. and G.M. analyzed data; I.E.J.d.V. wrote the paper.

This research was funded by a postdoctoral fellowship awarded by Fondazione Caritro to I.E.J.d.V.

The authors declare no competing financial interests.

Correspondence should be addressed to Ingmar E. J. de Vries at i.e.j.de.vries@gmail.com or Daniel Baldauf at daniel.baldauf@unitn.it.

https://doi.org/10.1523/JNEUROSCI.0583-21.2021

One candidate mechanism is oscillatory cortical activity in the alpha frequency range (8–14 Hz). The power and phase of cortical alpha oscillations modulate the firing of underlying neuronal populations and predict subsequent sensory discrimination (Haegens et al., 2011b, 2015). Important here, alpha activity is functionally modulated in anticipation of auditory selection (Weisz et al., 2011). For example, prospective cues indicating lateralized auditory targets result in alpha lateralization in auditory and parietal cortices (Banerjee et al., 2011; Müller and Weisz, 2012; Ahveninen et al., 2013; Frey et al., 2014), paralleling a well described effect in visual attention (Sauseng et al., 2005; Thut et al., 2006; Bagherzadeh et al., 2020). These effects are not merely epiphenomenal but have a functional role in controlling attention and perception. For instance, alpha modulations predict attentional gain of the cortical representation (Kerlin et al., 2010) and subsequent behavioral performance on auditory tasks (Obleser and Weisz, 2012; Leske et al., 2015; Herrmann et al., 2016; Wöstmann et al., 2019a,b). Additionally, alpha transcranial alternating current stimulation modulates target recall (Wöstmann et al., 2018), and phantom sounds (i.e., tinnitus) can be reduced by alpha repetitive transcranial magnetic stimulation (Müller et al., 2013) or neurofeedback (Hartmann et al., 2014). Interestingly, alpha modulations seem to have a dual role, such that alpha suppression facilitates processing of relevant stimuli (Leske et al., 2015; Griffiths et al., 2019), whereas alpha enhancement attenuates the processing of distracting stimuli (Strauß et al., 2014; Wöstmann et al., 2017). Careful independent manipulation of the spatial characteristics of targets and distractors has revealed these processes to act simultaneously (Wöstmann et al., 2019a). Similarly, much research on auditory attention has involved manipulating and cueing single features in relatively simple stimuli (Hill and Miller, 2010; Ahveninen et al., 2013; Ding and Simon, 2013). However, naturalistic auditory scenes usually comprise a complex mixture of auditory signals that overlap in their feature content, and that originate from hard-to-distinguish spatial sources. Rather than using crude differences in feature information, we depend on object-based auditory attention (Griffiths and Warren, 2004). It remains unclear how the brain prepares for an anticipated auditory object of interest within an ecologically valid auditory scene.

Here we investigated how anticipatory alpha oscillations are functionally involved in object-based auditory attention. We cued subjects in which of two spatially and temporally overlapping naturalistic auditory streams a repetition of an auditory object was most likely to appear. Using traditional univariate methods, it is difficult to distinguish neural mechanisms of target selection and distractor suppression in such a complex naturalistic sound scene. We therefore adopted a multivariate approach and trained a linear classifier on MEG source-reconstructed oscillatory activity to dissociate which stream was attended in anticipation of the repetition. We hypothesized anticipatory alpha oscillations, indexing top-down attentional control, to be timely related to the subsequent identification of the repeated auditory object.

Materials and Methods

Subjects

Fifteen healthy volunteers (mean age, 28 ± 3 years; 9 females) participated in the experiment for monetary compensation. All subjects had normal or correct-to-normal vision and were tested for a balanced left–right hearing perception using a sample of the stimuli from the main experiment. Subjects were naive with respect to the purpose of the study. All experimental procedures were performed in accordance with the Declaration of Helsinki and were positively reviewed by the Ethical Committee of the University of Trento. Written informed consent was obtained. The entire session including preparation lasted ~2 h, of which 1.2 h were spent in the MEG scanner.

Experimental design

The experiment was created using the Psychophysics Toolbox (version 3.72; RRID:SCR_002881) in MATLAB (version 2012b; MathWorks; RRID:SCR_001622). Subjects performed an auditory attentional cueing paradigm (Fig. 1*A*; same design and stimuli as experiment 1 in the study by Marinato and Baldauf (2019), but adapted for MEG). For clarity and completeness, the full experimental design and parameters are described here. Each trial started with a fixation cross (1–2 s, randomly jittered), consecutively followed by a visual cue (0.5 s) and a delay period (0.5–0.75 s, randomly jittered), with the auditory scene consisting of two overlapping auditory signals (i.e., a "speech" and an "environment" signal; 5 s), and an additional fixation cross (1.5 s) to allow response times (RTs) to extend beyond the auditory stimulation. The visual cue indicated in which of the two overlapping auditory signals a to-be-detected repetition would occur. It consisted of the capital letter "S" if the repetition would occur in the speech signal, the capital letter "E" if it would occur in the environment signal, or both letters in case the repetition could occur in either of the two signals (i.e., the neutral cue condition). The cue was valid in 70% of trials, was invalid in 20% of trials, and was neutral in the remaining 10% of trials. Subjects were instructed to pay attention to the cued stream and to respond with the right index finger on a button press when they heard a repetition in one of the auditory signals. Speed and accuracy were equally emphasized.

Subjects performed one practice block consisting of 100 trials. During the practice block, we presented only one of the two auditory signals, such that subjects could more easily understand what a repetition sounded like. This practice phase lasted for ~17 min. Next, subjects performed three experimental blocks consisting of 100 trials each, during which we always presented the auditory scenes consisting of overlapping speech and environmental signals. The factors of cue validity and position of the repetition (i.e., the speech or the environmental signal) were randomly mixed within blocks.

Auditory stimuli

The auditory scenes (Fig. 1*B*) consisted of the following two overlapping signals: a conversation (i.e., Speech) and an environmental sound (i.e., Environment). Speech signals comprised 5 s segments extracted from newscast recordings of various foreign languages. Importantly, subjects were unfamiliar with these languages. The environmental sounds consisted of field recordings of public places such as airports, streets and restaurants. We dynamically modulated the envelope of the environmental sounds using envelopes randomly extracted from the speech signals to make the two streams as comparable as possible in terms of low-level features (i.e., the envelop). Furthermore, to control for spatial confounds we converted the two streams in mono by averaging the stereo channels together and presenting the single resulting signal diotically (i.e., simultaneously to both ears). The repetition consisted of a randomly sampled 0.75 s segment extracted from the auditory stimuli, which were inserted twice in sequence into the corresponding signal. The length of the repetition was chosen to approximately correspond to a functional unit (i.e., an auditory object) like a typical acoustic event in environmental sounds or a couple of syllables/words in normal speech. Crucially, given the complexity of the stimuli and the difficulty of the task, this task was very difficult to solve by attending low-level features, but was readily solvable by processing the objects fully using object-based attention. Specifically, in both streams the individual auditory features were short lived and variable over the time course of a 750-ms-long "repetition segment" (possibly containing several auditory objects, such as different words in the speech signal and cars, coffee machines, footsteps, glasses clinking together while a waiter cleans up a table, doors opening/closing, chairs being pushed in the environment signal). Some auditory objects or single acoustic features would naturally reoccur almost identically or very similarly, outside of the context of the repetition segment. Therefore,

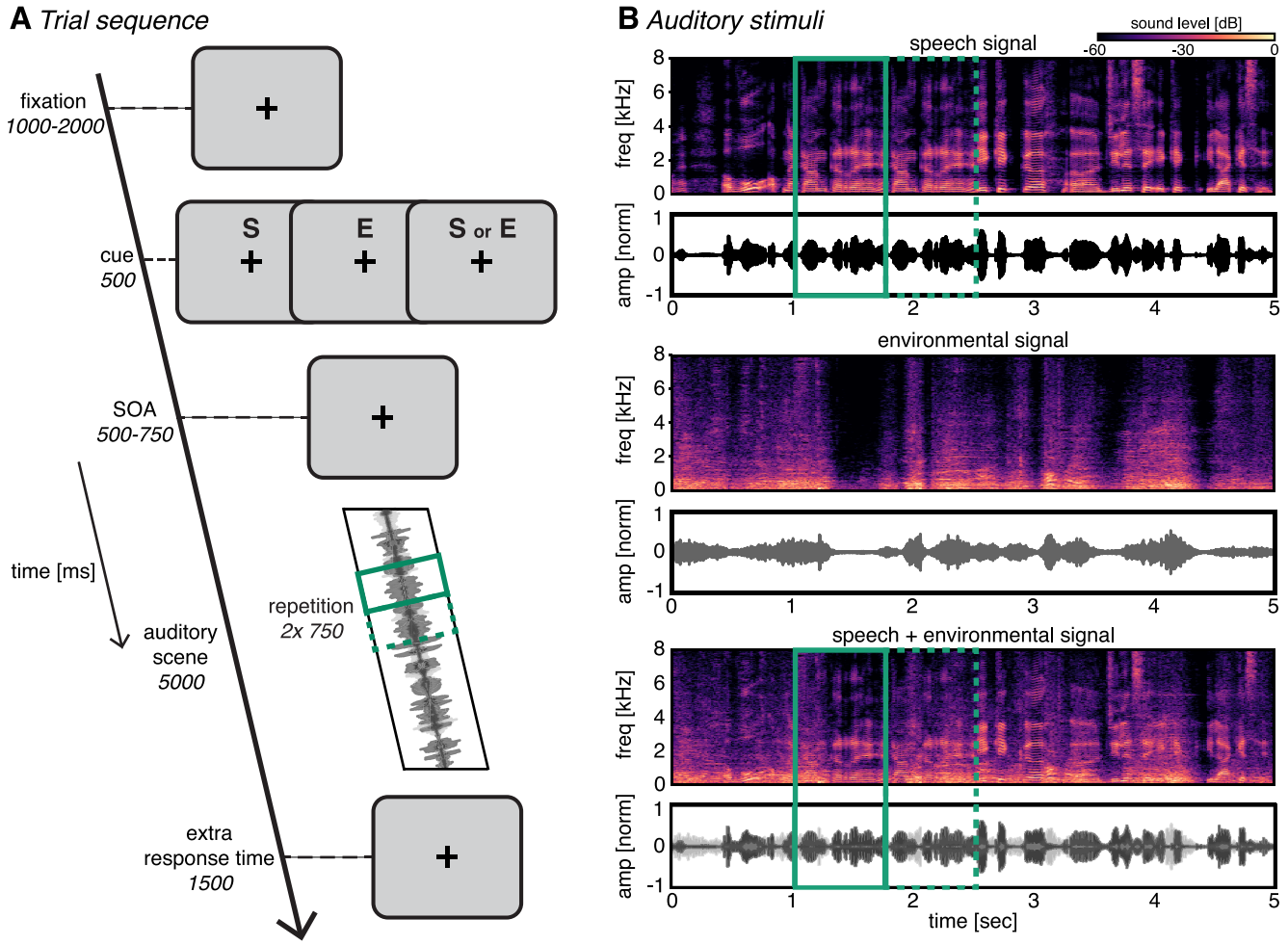

Figure 1. Task design. ***A***, Trial sequence. Subjects were presented with an auditory scene consisting of two overlapping streams. They were instructed to detect and respond to a repetition in one of the two streams. A cue indicated whether the repetition would appear in the speech stream (S), the environmental sound stream (E), or either stream (S or E). ***B***, Example auditory scene (bottom), consisting of a speech signal (top) and an environmental signal (middle). Time–frequency plots represent the spectral content of the streams; time series represent the normalized amplitude. Full and dashed green outlines indicate the first and second instance of an example repetition, respectively. Note that the temporal location of the repetition within the sound scene was randomized across trials.

participants needed to recognize and categorize these objects, put them in context, and ultimately have a full representation of the repeated segment within the sound scene to recognize the repetition. Linear ramping and cross-fading algorithms were applied to avoid cutting artifacts and to render the transition between segments unnoticeable.

MEG recording and preprocessing

Whole-head MEG recordings were obtained at a sampling rate of 1000 Hz using a 306-channel (204 first-order planar gradiometers, 102 magnetometers) VectorView MEG system (Neuromag, Elekta) in a two-layer magnetically shielded room (AK3B, Vacuum Schmelze). A low-pass antialiasing filter at 330 Hz and a high-pass filter at 0.1 Hz were applied online. Before the MEG recording, we digitized the individual head shape with an electromagnetic position and orientation monitoring system (FASTRAK, Polhemus) using the positions of three anatomic landmarks (nasion and left and right preauricular points), five head position indicator coils, and ~300 additional points evenly spread on the subject's head. Landmarks and head-position induction coils were digitized twice to ensure a localization error of <1 mm. To coregister the head position in the MEG helmet with anatomic scans for source reconstruction, we acquired the head positions at the start of each run by passing small currents through the coils. Upon detection of the target repetition, subjects responded by pressing with their right index finger on an MEG-compatible button press, which was together with all other hardware connected to a DataPixx input/output hub to deliver visual cues and sound stimuli and to collect button presses in a critical real-time manner (VPixx Technologies; RRID:SCR_009648).

MEG data were preprocessed offline using a combination of the Brainstorm (Tadel et al., 2011; RRID:SCR_001761) and Fieldtrip (Oostenveld et al., 2011; RRID:SCR_004849) toolboxes in MATLAB, as well as custom-written MATLAB scripts, following general standards for MEG preprocessing (Tadel et al., 2019). Continuous data from each run were visually inspected for noisy sensors and system-related artifacts (e.g., SQUID jumps), and a maximum of 12 noisy sensors were removed from further analyses. Next, we applied the Neuromag MaxFilter implementation of Signal Source Separation (SSS; Taulu and Simola, 2006) for removing external noise to each individual run, and an extended Infomax independent component analysis (ICA; Lee et al., 1999) for removing components capturing blinks and eye movements. Note that the maximum number of ICA components to be extracted was determined by the residual degrees of freedom after SSS rank reduction. We removed an average of 2.2 ± 0.6 ICA components. Continuous data were then segmented in epochs from −2 to 7 s locked to the onset of the auditory stimulus. Each epoch was visually inspected, and those containing artifacts not cleaned by previous preprocessing steps were discarded from further analyses. All data cleaning steps combined resulted in an average of 19% of all trials rejected per subject.

Source reconstruction

For the construction of 3D forward models for MEG source reconstruction, we first acquired previously recorded anatomic 3D images, if

available (nine subjects), and used the standard brain from FreeSurfer (RRID:SCR_001847) for the remaining subjects. Individual anatomic images were obtained using a 4 T magnetic resonance imaging (MRI) scanner (Bruker Biospin), with an eight-channel birdcage head coil (magnetization-prepared rapid gradient echo, $1 \times 1 \times 1$ mm). The anatomic scans were then 3D reconstructed using FreeSurfer default settings (Dale et al., 1999; Fischl et al., 1999). For the six subjects without individual anatomy, the standard FreeSurfer anatomy was warped to the subject's head volume as estimated from the digitized head shape. Next, head models were computed by coregistering the subjects head shape with the reconstructed MRI brain volumes using FreeSurfer default settings and using overlapping spheres. We then performed source reconstruction using minimum-norm estimates for a source space of 15,000 vertices (Hämäläinen and Ilmoniemi, 1994), as implemented in Brainstorm. To allow for intersubject comparisons, the averaged source maps were normalized with respect to a 200 ms baseline window (z-scores). Source activity was estimated for each run separately, after which we combined data from different runs. Last, each individual source space was projected to a standard FreeSurfer brain and parcellated according to the recently developed cortical atlas from the Human Connectome Project (HCP), which provides the most precise insights into the structural and functional organization of the human cortex to date (Glasser et al., 2016). This parcellation is based on a multimodal atlas of the human brain obtained by combining structural, diffusion, functional, and resting-state MRI data from 210 healthy young individuals and identifies 180 regions of interest (ROIs) per hemisphere. We averaged the signals from all vertices within an ROI to obtain a total of 360 time series of estimated cortical activity.

Morlet wavelet convolution

The 360 ROI time series were decomposed into time frequency representations with Morlet wavelet convolution using a custom-written MATLAB script, for frequencies ranging from 1 to 40 Hz in 25 logarithmically spaced steps. A Gaussian ($e^{-t^2/2s^2}$, where s is the width of the Gaussian) was multiplied with 25 sine waves ($e^{i2\pi ft}$, where i is the complex operator, f is frequency, and t is time) to create complex Morlet wavelets. The width was set as $s = \delta/(2\pi f)$, where δ represents the number of cycles of each wavelet, logarithmically spaced between 3 and 12 to have a good trade-off between temporal and frequency precision (Cohen, 2014). We applied frequency domain convolution by multiplying the ROI signals with the Morlet wavelets after applying the fast Fourier transform (FFT) to each. This step was followed by a conversion back to the time domain using the inverse FFT. The squared magnitude of these complex signals was taken at each time point and each frequency to acquire power (i.e., $[\text{real}(Z_t)^2 + \text{imag}(Z_t)^2]$), after which power was downsampled to 50 Hz to reduce computation time. Last, before multivariate pattern classification, the power at each time point, frequency, and ROI was z-normalized across trials (Newman and Norman, 2010; Jafarpour et al., 2013).

Multivariate pattern classification

Our main analysis involved a backward-decoding classification algorithm (linear discriminant analysis) on the time-frequency-decomposed power, with all 360 ROIs as features and the "attend speech" and "attend environment" labels as classes. This analysis tests whether a linear classifier can learn to dissociate between attending to either the speech signal or the environmental signal from cortical patterns of oscillatory power modulations. The complete classification analysis was performed separately on the data of each individual subject. We used a linear classifier as implemented in the Amsterdam Decoding and Modeling toolbox (ADAM; Fahrenfort et al., 2018), an open source, script-based toolbox in MATLAB for backward-decoding and forward-encoding modeling of EEG/MEG data. Note that we replaced the standard time-frequency decomposition in the toolbox with the custom-written Morlet wavelet convolution described above (de Vries et al., 2019; van Driel et al., 2019), and based on Cohen (2014), as this arguably provides a better trade-off between temporal and spectral precision. We applied the following 10-fold cross-validation procedure: first, the trial order was randomized, and trials were partitioned in 10 equal-sized folds; next, a leave-one-out procedure was used in which the classifier was trained on 9 folds and tested on the remaining fold. This procedure was repeated 10 times until each fold was used exactly once for testing, after which classifier performance was averaged over folds. We applied between-class balancing using oversampling to ensure that the classifier would not develop a bias for the overrepresented class during training. Because the design was balanced in terms of trial counts for attention conditions (attend speech vs attend environment), between-class balancing was only necessary to eliminate small imbalances because of trials rejected during data cleaning. Crucially, by combining all trials in which different speech exemplars were cued in the attend speech class, and all trials in which the different environment exemplars were cued in the Attend Environment class, our classification analysis is oblivious to the exact auditory signals themselves. This is in contrast to speech-tracking approaches, in which the exact auditory signals (or their temporal envelope) are tracked. This analysis therefore better captures high-level attentional control mechanisms, and it is an elegant solution to decode the auditory attentional state without the need for information from the input audio signal.

We adopted the area under the curve (AUC) as a measure of classifier performance, with the curve being the receiver-operating curve of the cumulative probabilities that the classifier assigns to trials coming from the same class (true positives) against the cumulative probabilities that the classifier assigns to trials that come from the other class (false positives). An AUC value of 0.5 means chance-level classification performance. Instead of averaging across binary decisions about class membership of individual trials (as with standard classification accuracy), the AUC incorporates the level of confidence (i.e., the distance from the decision boundary) that the classifier has about class membership of individual trials. The AUC is considered a sensitive, nonparametric, and criterion-free measure of classification performance (Hand and Till, 2001).

Next, to investigate the cortical distribution of neural activity underlying significant classification, we computed cortical maps by multiplying classifier weights of all ROIs with the covariance matrix of the data across ROIs (Haufe et al., 2014), as implemented in ADAM (Fahrenfort et al., 2018). Note that this is the covariance matrix of the 360 source-reconstructed ROI signals, not the covariance matrix of the initial sensor data used for MNE source reconstruction. An important caveat when looking at classifier weights is that a certain ROI might have a high classifier weight because it helped remove noise that was not task related and therefore helped the classifier to perform better (Haufe et al., 2014; Fahrenfort et al., 2018). In contrast, the transformation procedure used here generates activation patterns that return the mass-univariate difference between the compared conditions, which, unlike classifier weights, can be interpreted as neural sources. To make activation values comparable between subjects and to allow for averaging, the individual subject activation patterns were spatially normalized by subtracting the mean across ROIs and dividing by the SD across ROIs (i.e., z-scored; Haufe et al., 2014; Fahrenfort et al., 2017, 2018). Next, since the directionality of these activity maps is arbitrary depending on condition order, we took the absolute for plotting only to highlight the magnitude of involvement of different ROIs. We plotted these cortical maps on the FreeSurfer standard brain, by coloring the cortical areas as defined by the 360 ROIs from the HCP atlas that formed the features for our classifier in Brainstorm (Figs. 3*D*, 4*E*). Note that we did not perform any statistical test on these activation patterns, because the strength of multivariate pattern analysis (MVPA) is the fact that a classifier can use any bit of information contained in any of the ROIs to separate classes. This does not mean that all of the individual ROI weights are significant or that they have to be. Highlighting significant ROIs might give the impression that no other ROI contained task-relevant information, which would be misleading.

Feature, trial, and time selection

In an initial analysis, we used the 360 ROIs as features and performed the classification analysis on each frequency and each time point separately, thus yielding classification performance over time and frequency (Figs. 3*A*, 4*A*). Next, because we hypothesized an active role of alpha oscillations in auditory attention, we a priori selected the alpha band for a more sensitive classification analysis. Here, instead of using 360 ROIs as features and performing the classification analysis on each frequency separately, we used the individual frequencies within the alpha band as an additional feature dimension for the classifier (Fuentemilla et al.,

2010; Jafarpour et al., 2013). That is, our alpha band (8–14 Hz) contained four frequencies, which, combined with 360 ROIs, resulted in 1440 features that were fed into the classifier. This approach has a very important advantage. The alpha peak frequency varies highly both between subjects, within subjects in different brain regions, and within subjects during different brain states (e.g., rest vs passive viewing vs demanding task; Haegens et al., 2014). Feeding all frequencies within a wide alpha band (i.e., 8–14 Hz; de Vries et al., 2020) as features into the classifier allows the classification algorithm to find the most information within the alpha band during our cognitive process of interest in a data-driven manner, and on an individual subject basis, thus obviating the need for the selection of peak frequency. It thus allows for individual differences in peak frequency regarding information content for the classifier. Furthermore, it is even sensitive to this peak frequency of information content being different in different regions within the same subject (e.g., 8 Hz in frontal cortex can be independently weighted from 12 Hz in auditory cortex). Additionally, in the repetition onset-locked analysis, we observed significant classification in the delta to low-theta band (i.e., 2–5 Hz; Fig. 4*A*). Therefore, as an exploratory analysis, we performed the same above-mentioned more sensitive analysis that uses all the frequencies within this band as features for the classifier.

For the auditory stimulus onset-locked analysis, we trained and tested a classifier for each time point from −0.5 to 3 s surrounding stimulus onset. Importantly, at each time point we selected only those trials in which the repetition was not presented yet (Table 1, complete overview of trial counts included in each of our analyses). This was to ensure that classification was not driven by the perception of the actual repetition, but rather by endogenous attentional orienting toward one of the two overlapping auditory streams in anticipation of the repetition. Note that this trial selection procedure resulted in a gradual decline in trial numbers being included in the classification analysis the further we moved in time after stimulus onset. This explains why classifier performance becomes noisier further in time in the stimulus-locked analyses (Fig. 3). The end point of $t = 3$ s post-stimulus onset was selected because this was the median (and approximately the mean) of all possible repetition onsets, and trial counts became relatively low for some of our classification analyses after that time point. Furthermore, because we were interested in the effect of endogenous attentional orienting driven by the cue, we excluded neutral cue trials. In a separate analysis, we trained and tested separate classifiers on correct or error trials only, as an additional investigation of the relevance of frequency-specific classification for actual behavioral performance. For this analysis, an error was defined as no response, a response before repetition onset, or within 300 ms of repetition onset, as this was unlikely to be driven by an actual detection of the repetition. The trimming of response time data at a cutoff of 300 ms is a common procedure and is based on the argument that the processing chain from target detection to response execution takes more time (Ratcliff, 1993; van Moorselaar et al., 2014; de Vries et al., 2017). In the current experiment in which repetition detection does not rely on detecting a single feature, but instead on the repetition of a pattern over an extended period of time (i.e., an auditory object), it is even less likely that a response faster than 300 ms after repetition onset is driven by an actual detection of the repetition. In any case, the bulk of error trials (i.e., 96%) consisted of trials in which no response was given, or trials in which a response was given before repetition onset, and thus this particular cutoff selection will have a negligible effect. Because there were more correct trials than error trials, any difference in classification performance could be driven by a difference in trial count (and thus in signal-to-noise ratio). Therefore, we also performed the classification analysis on a randomly selected subset of correct trials that matched the individual subject's count of error trials.

For the repetition-locked analyses, trials were realigned in time to the onset of the repetition, and classification analyses were applied to each time point from −0.5 to 1.5 s surrounding this onset. For this analysis, we were interested in the neurocognitive processes underlying the detection of the repetition, rather than anticipatory attention driven by the cue. Therefore, for this analysis we did include neutral cue trials. Furthermore, in this analysis only, for invalid cue trials the speech versus environment event codes were swapped, such that we classified in which stream the repetition was actually presented, rather than which stream was cued.

Table 1. Trial counts per classification analysis

	Stimulus locked	Repetition onset locked
All trials	252 ± 18 (from −0.5 to 1.76 s) 125 ± 10 (at 3 s)	280 ± 20
Correct response trials only	145 ± 18 (from −0.5 to 1.76 s) 68 ± 9 (at 3 s)	161 ± 20
Error response trials only	107 ± 19 (from −0.5 to 1.76 s) 56 ± 11 (at 3 s)	119 ± 22

Numbers indicate subject average ± SD. Note that neutral cue trials were excluded from the stimulus-locked analyses, while they were included in the repetition onset-locked analyses.

To address the question of whether the significant stimulus- and repetition-locked alpha classification effects observed here (Figs. 3*B*,*C*, 4*B*, *C*) reflect the same or different neurocognitive processes, we adopted a generalization-across-time (GAT) approach (King and Dehaene, 2014). That is, a classifier was trained on data in the prerepetition interval locked to stimulus onset, and tested on data in the postrepetition interval locked to repetition onset. Because the two intervals are locked to a different event, it is impossible to create a typical GAT matrix in which the diagonal represents training and testing at the same time points. Instead, the horizontal and vertical time axes represent nonoverlapping time points (Fig. 5, top row, GAT plots). The exact same trial selection procedures were performed as in the main stimulus-locked and repetition-locked analyses. Note that for the stimulus-locked training data, at each time point only trials were included in which the repetition was not presented yet, while for the repetition-locked testing data all trials were included. The repetition-locked testing data were only analyzed from repetition-onset onward (i.e., $t = 0$). In other words, there were no overlapping time points in the training and testing data, preventing double-dipping, and eliminating the need for n-fold cross-validation. However, because of temporal smearing inherent to wavelet convolution, a negligibly small artificial increase in classification accuracy can be observed around repetition onset (Fig. 5, bottom right corner in GAT plots, beginning of the curve plots). Additionally, to increase sensitivity we averaged over the stimulus-locked training time within the interval of significant classification in our main stimulus-locked analysis at $p < 0.01$ (Figs. 3*B*,*C*, significant intervals, 5, bottom row, resulting plots).

Statistical analysis

Behavior was analyzed with two repeated-measures ANOVAs for both RT and accuracy data using SPSS (version 21.0.0.0; RRID:SCR_002865), with the within-subject factor cue validity (valid, neutral, and invalid). We used the Greenhouse–Geisser correction for violations of sphericity, and pairwise comparisons were Bonferroni corrected for multiple comparisons. Effect sizes are reported as partial η squared (η^2) for ANOVAs, and Cohen's d for pairwise comparisons. Because our statistical tests of classifier performance over time (and frequency) involved many comparisons (each time-frequency point), we performed group-level nonparametric permutation testing with cluster-based correction for multiple comparisons, which controls for the autocorrelation over time and frequency (Maris and Oostenveld, 2007). First, for every time (frequency) point, we computed t values for the AUC deviation from chance (i.e., 0.5) and set a threshold at a certain p value (≤ 0.05; see Results), which resulted in clusters of significant time (frequency) points. Next, for each time (frequency) point, we randomly shuffled the sign of the AUC deviation from chance across subjects over 2000 iterations, and in each iteration performed a t test on the shuffled data for the AUC deviation from chance. In each iteration, we computed the size of the largest time (frequency) cluster of significant t values, which resulted in a null distribution of maximum cluster sizes under randomly shuffled data. The sizes of the significant time (frequency) clusters in the observed data were compared with this null distribution using a threshold corresponding to the p value used for the t tests (e.g., 99th percentile for $p < 0.01$). We constrained the statistical tests from −0.5 to 3 s surrounding auditory stimulus onset for the stimulus-locked analyses, and

from −0.5 to 2 s surrounding repetition onset for the repetition-locked analyses.

Exploratory correlation AUC–behavior

To further investigate the behavioral relevance of significant classification of anticipatory auditory attention from alpha oscillations, we tested how classification performance related to accuracy and reaction time on the response to the repetition. First, we selected the peak of alpha power classification performance (i.e., 1660–1880 ms, the time window in which AUC was significant at $p < 0.01$) in the analysis of correct trials only (Fig. 3*C*, green line) and averaged for each subject the AUC values over that time window. Next, we performed two across-subject Spearman rank correlation analyses. Between averaged AUC values and trial average reaction times (again for correct trials only, as error trials were defined by no response, a response before repetition onset, or a response within 300 ms of repetition onset). And between averaged AUC values and a subject's average accuracy (i.e., percentage of trials defined as correct response). Using this procedure, we thus linked significant classification of auditory attention from alpha oscillations to subsequent detection performance of the auditory repetition. Note that we did not a priori plan these correlation analyses, and that $N = 15$ is on the low side for a correlation analysis. These results should therefore be considered an exploratory addendum to better characterize the robust findings that emerge from the classification analysis and should enrich our understanding of the data. The combination of this correlation analysis and the separate classification analyses for correct and error trials provide converging support for the functional role of alpha oscillations in auditory object-based attention.

Control analyses

Fourier transforming of any signal in which events take place (e.g., our stimulus and repetition onsets), might lead to by-products in the frequency domain (Cohen, 2014). As such, something that appears as a peak in the spectral domain can reflect a stimulus-evoked response [as traditionally captured by the event-related field (ERF)], rather than a modulation in ongoing oscillatory activity. To ascertain that true oscillatory activity, rather than the stimulus-evoked response, drove the significant classification results, we ran two control analyses. First, for our main findings we performed the exact same analyses as described above, now performed on the time-resolved signal before applying wavelet convolution (i.e., the single-trial broadband signal), but after preprocessing and source reconstruction (Fig. 6*A*,*B*, left column). Next, we reran the same analysis, but now on non-phase-locked (i.e., induced) power rather than total power (Fig. 6*A*,*B*, middle, right columns). We computed non-phase-locked power by subtracting the average over trials (i.e., the ERF) from the raw data on each single trial, before applying wavelet convolution (Cohen, 2014). This step removes the stimulus-evoked component from total power and, thus, extracts modulations in ongoing oscillations that are not locked to the stimulus.

For the main stimulus-locked analysis, at each time point we selected only those trials in which the repetition had not taken place yet. However, wavelet convolution inevitably results in temporal smearing. While a wavelet has a Gaussian shape, thus having the strongest weighting at the center of the wavelet, any effects observed before repetition onset could potentially be influenced by repetition-induced effects that are smeared back in time. We performed an additional control analysis to exclude this possibility. The longest wavelet within our alpha range is at 8.6 Hz with 8.25 cycles (i.e., a wavelet 0.96 s long). This means that at a certain time point there is smearing over a temporal window of 0.48 s in either direction. We thus performed the exact same analysis as our main analyses using all trials or correct trials only (Fig. 3*B*,*C*, respectively), but now at each time step we selected only those trials in which the repetition would not take place within the next 0.48 s (Fig. 6*C*).

Given that there was a difference in repetition detection performance between attend speech (65% detection rate) and attend environment (53% detection rate), one potential confound is that significant classification accuracy could be driven by a difference in task difficulty between the two classes. Therefore, we performed two control analyses on our main results of significant stimulus-locked alpha classification (Fig. 3*B*, *C*), for which we operationalized task difficulty as behavioral accuracy (i.e., detection rate). First, we performed the exact same analysis as in our main analysis using all trials (Fig. 3*B*), but now after equalizing the ratio of correct trials to error trials between the two classes (Fig. 6*D*). Note that our analyses using only correct trials or only error trials (Fig. 3*C*) are inherently balanced concerning performance. Second, we correlated across subjects the performance difference between attend speech and attend environment trials with decoding accuracy averaged over the interval of significant classification ($p < 0.01$; Fig. 3*B*, gray horizontal bar). If a difference in task difficulty would be the main driver of classification accuracy, one would expect a significant positive correlation.

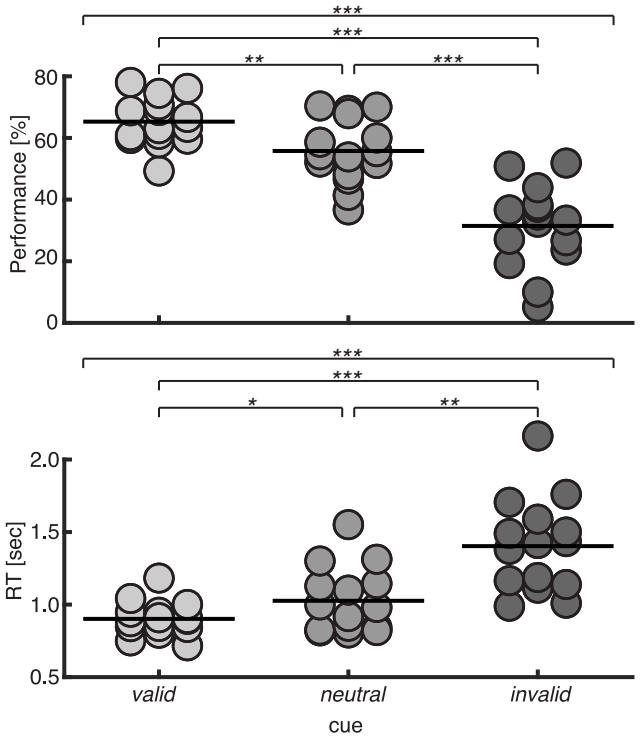

Figure 2. Behavioral results. Dots represent single-subject data (percentage correct and trial-averaged correct RT in top and bottom panels, respectively). Horizontal thick lines represent the group mean. The top thin horizontal line in each panel represents the result of a repeated-measures ANOVA, and the other thin horizontal lines represent pairwise comparisons: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Data availability

All custom-written analysis scripts, other than the functions implemented in the Brainstorm and ADAM toolboxes, are freely available at https://osf.io/efv4b/. Raw MEG data will be shared on reasonable request.

Results

Behavior

As expected, the attentional cue had the desired behavioral effect (Fig. 2). That is, performance was best on valid cue trials (accuracy, 65 ± 8%; RT, 901 ± 116 ms), it was reduced on neutral cue trials (56 ± 10%, 1026 ± 226 ms), and was the worst on invalid cue trials (31 ± 13%, 1403 ± 320 ms). These differences resulted in a significant main effect of the cue on both accuracy and reaction time (accuracy: $F_{(1.5,20.4)} = 49.6$, $p < 0.001$, $\eta^2 = 0.78$; reaction time: $F_{(1.5,19.5)} = 22.8$, $p < 0.001$, $\eta^2 = 0.62$). *Post hoc* pairwise comparisons indicated a difference between each of the cue conditions on accuracy (valid vs neutral: $p = 0.002$, $d = 1.12$; neutral vs invalid: $p < 0.001$, $d = 1.59$; valid vs invalid: $p < 0.001$, $d = 2.16$) and on reaction time (valid vs neutral: $p = 0.048$, $d = 0.71$; neutral vs invalid: $p < 0.002$, $d = 1.12$; valid vs

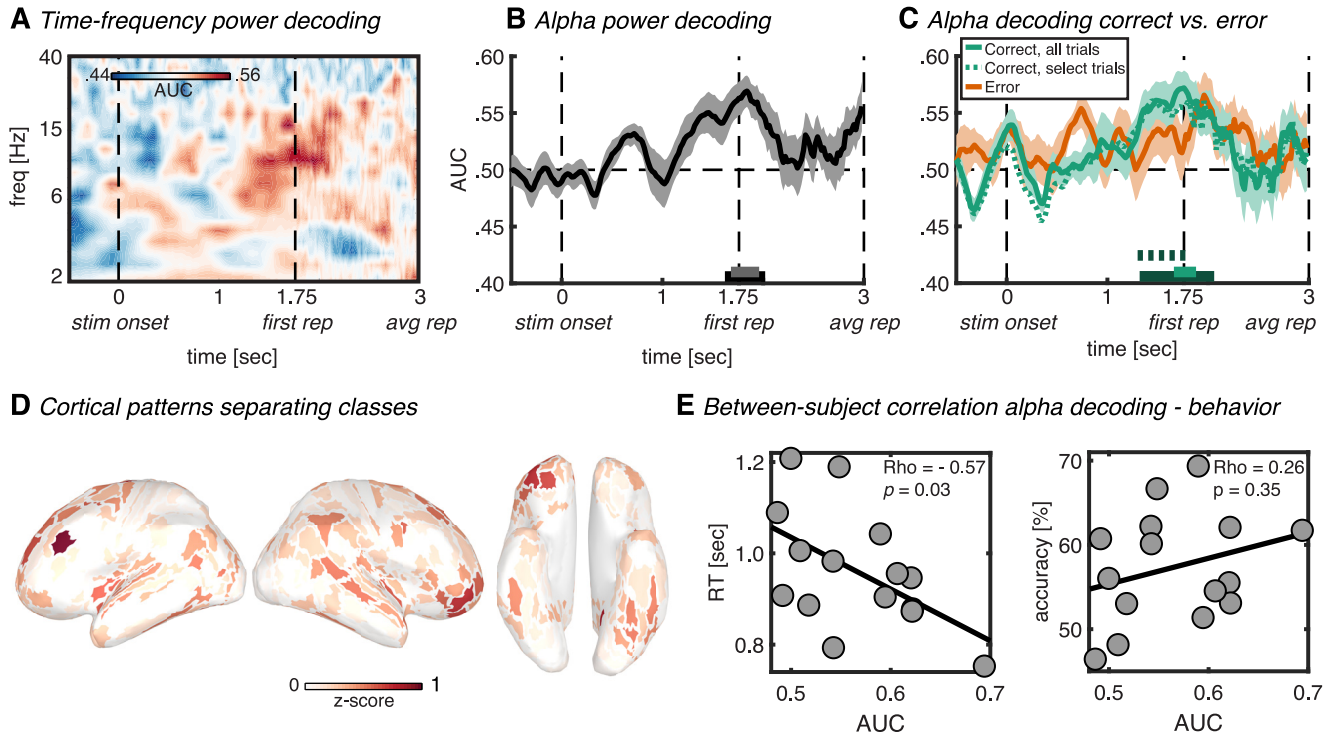

Figure 3. Auditory stimulus-locked decoding. ***A***, Time–frequency map of classifier accuracy (AUC). ***B***, Classifier accuracy plotted over time for the alpha band (8–14 Hz). The horizontal bars on the *x*-axis indicate significant classification after cluster correction at $p < 0.05$ (black) or $p < 0.01$ (gray). ***C***, Alpha power classifier accuracy as in ***B***, but with separate classification analysis for correct (green) and error (orange) trials. The dotted green line signifies classification using a random selection of correct trials equal to the amount of error trials per individual subject. Green horizontal bars (full and dotted) denote significant cluster-corrected classification (dark green, $p < 0.05$; light green, $p < 0.01$). Thick lines and shaded areas in ***B*** and ***C*** denote subject mean and SEM, respectively. ***D***, Cortical map of activation pattern separating classes in the significant ($p < 0.01$) interval indicated in ***C*** by the light green horizontal bar. Maps were *z*-scored before averaging, and absolute values were taken to highlight contributing ROIs. Left lateral, right lateral, and ventral views are shown. ***E***, Between-subject correlation between alpha power classification accuracy (AUC) for correct trials only in the significant ($p < 0.01$) interval indicated in ***C*** by the light green horizontal bar, and reaction time (left) or accuracy (right). Dots represent single-subject values, whereas diagonal lines represent least-square fits. *stim onset*, Auditory stimulus onset; *first rep*, time of first repetition averaged over subjects; *avg rep*, time of repetition averaged over subjects and trials.

invalid: $p < 0.001$, $d = 1.41$). This replicates previous behavioral findings from our laboratory using the same paradigm (Marinato and Baldauf, 2019).

Stimulus-locked multivariate pattern classification

Our main analysis investigated whether we could classify from the cortical pattern of time-frequency-decomposed oscillatory activity, which of the two overlapping auditory streams subjects attended. Specifically, at each time point we trained a linear discriminant classifier to dissociate between the attend environment and attend speech conditions, using frequency-specific power at all 360 scouts as features (Fig. 3*A*). Given that, on the basis of previous studies, we had an a priori hypothesis about the alpha frequency band, we performed the same classification analysis, but now using the different frequencies within the alpha band at each scout as separate features, thus creating a total of 1440 features (Fuentemilla et al., 2010; Jafarpour et al., 2013; Fig. 3*B*).

In line with our expectations, this analysis revealed auditory attention-sensitive information in the alpha band of the source-reconstructed MEG signal (from 1.62 to 2.02 s and from 1.68 to 1.96 s relative to stimulus onset, cluster-corrected at $p < 0.05$ and $p < 0.01$, respectively; Fig. 3*B*). The initial classification analysis over the whole frequency range confirmed the peak of classification to be located in the alpha frequency range (Fig. 3*A*). Interestingly, alpha classification time series show a gradual increase in classification accuracy with a peak at around the onset of the first repetition (Fig. 3*B*), which suggests that over many trials subjects implicitly learned to anticipate the timing of repetition onsets and thus knew when to focus their attention on the correct stream. Before that time, no repetitions were expected yet. This is in line with research on temporal attention (Nobre and Van Ede, 2018), in which it has been repeatedly observed across sensory modalities that preparatory alpha modulations are transient and that they peak just before the expected target onset, only to disappear again immediately after (Praamstra et al., 2006; Rohenkohl and Nobre, 2011; Van Ede et al., 2011; Zanto et al., 2011; Wöstmann et al., 2021). The transient nature of the peak suggests that it reflects attentional selection of the relevant auditory input, rather than a steady-state signature of attentional tracking. Once selected, the diagnostic alpha signature disappears. Together, these results thus support a role for alpha oscillations in object-based auditory attention, and suggest that they are flexibly focused in time according to temporal expectations about the necessity to pay attention (i.e., at anticipated repetition onset).

To further investigate the functional significance of the observed attention classification from alpha oscillations, we performed two exploratory follow-up analyses. First, we conducted the same classification analysis as for Figure 3*B*, but now separately for correct and error trials, which were defined as trials in which the subject either detected or did not detect the repetition, respectively (Table 1, number of trials included in each analysis). Interestingly, we observed significant alpha power classification for correct trials (from 1.32 to 2.06 s and from 1.66 to 1.88 s relative to stimulus onset, cluster-corrected at $p < 0.05$ and $p < 0.01$, respectively; Fig. 3*C*, full green line), but not for error trials (Fig. 3*C*, orange line). There were more correct trials compared with

error trials, which could potentially contribute to higher classification accuracy for correct trials. To exclude this possibility, we performed the analysis on correct trials again, but now on a random selection of correct trials up to the same amount as error trials per individual subject. This confirmed significant classification for correct trials from 1.30 to 1.82 s ($p < 0.05$, cluster corrected) relative to stimulus onset (Fig. 3*C*, dotted green line). A direct comparison between correct and error trials resulted in a significant difference ($p = 0.039$) at a single time point within the interval of significant classification (1.74 s relative to stimulus onset; i.e., right before the first repetition onset). However, this single time point did not survive cluster correction for multiple comparisons and should therefore be interpreted with care. As a second follow-up analysis, we correlated, across subjects, the classification accuracy (i.e., AUC) averaged over the significant interval at $p < 0.01$, with the reaction time and accuracy data (Fig. 3*E*). We found a significant negative correlation with reaction time (Rho = -0.57, $p = 0.03$), which suggests that subjects who were better prepared (i.e., attended better to the correct stream right before the repetition, as indicated by higher alpha classification accuracy) responded faster to the repetition. We did not observe such an effect for accuracy (Rho = 0.26, $p = 0.35$). Albeit speculative, these two results suggest a causal role for oscillatory alpha activity in object-based auditory attention since significant alpha classification earlier during the trial predicted subsequent behavioral performance.

The cortical patterns of forward-transformed classifier weights in the time window of significant classification (Fig. 3*D*) showed a distributed and complex contribution of multiple cortical areas. Corroborating earlier research on auditory attention, there was involvement of higher parts of early auditory cortex and auditory association cortex (Mesgarani and Chang, 2012; Lee et al., 2013; Puvvada and Simon, 2017; O'Sullivan et al., 2019), neighboring areas in insular and opercular cortex and the temporal–parietal–occipital junction (Bamiou et al., 2003; Vaden et al., 2013; Alho et al., 2015; Alavash et al., 2019), and several visual cortical areas (Cate et al., 2009; Vetter et al., 2014), among which are fusiform face area and parahippocampal place area (PPA; He et al., 2013; Bi et al., 2016; Bedny, 2017). Last, there was involvement of several frontal executive-control areas involved in scene navigation (Vann et al., 2009), the monitoring of expected events (Petrides, 2005), much like our current task demands. Together, this pattern confirms a complex involvement of multiple cortical regions in object-based auditory attention, with a prominence in areas involved in higher-level auditory cognition, executive-attention functions including the monitoring of expected events, and multimodal scene perception/navigation.

Repetition-locked multivariate pattern classification

In a second analysis pipeline, we realigned single-trial data to repetition onset after preprocessing and reran the time–frequency and classification analyses, as above, as this might shed more light on the neurocognitive processes involved in the detection of the repetition. Again, we found significant alpha power classification, but now more sustained from around the time of the average response (from 0.62 to 2.00 s and from 0.80 to 1.28 s relative to repetition onset; cluster corrected at $p < 0.05$ and $p < 0.01$, respectively; Fig. 4*A*,*B*). The follow-up classification analyses for correct and error trials separately confirmed the functional significance of alpha oscillations for auditory attention (Fig. 4*C*). That is, we observed an interval of significant alpha power classification for correct trials before the average response (from 0.36 to 0.66 s and from 0.38 to 0.64 s relative to repetition onset; cluster corrected at $p < 0.05$ and $p < 0.01$, respectively; Fig. 4*C*, full green line). Again, alpha power classification was not significant for error trials (orange line). A direct comparison between correct and error trials resulted in a significant difference from 0.32 to 0.60 s relative to repetition onset (cluster corrected at $p < 0.05$). Interestingly, the early interval of significant alpha power classification for correct trials falls within the time of the repetition presentation (which lasted 750 ms), thus supporting our interpretation of the stimulus-locked analyses for a functional role of alpha oscillations in successful detection of auditory objects.

After separating correct trials from error trials, we observed a second interval of significant classification from alpha power, after the average response (from 1.26 to 2.00 s, from 1.32 to 1.80 s, and from 1.38 to 1.56 s relative to repetition onset; cluster corrected at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively; Fig. 4*C*, green line), and an absence thereof in error trials (orange line). This second interval took place after the average response. Albeit speculative, it may be that after subjects successfully detected and responded to the repetition, they paid extra attention to that stream. Perhaps surprisingly, in the repetition-locked analysis we did not observe significant alpha decoding right before repetition onset, although we did observe it in the stimulus-locked analysis (Fig. 3). Note, however, that repetition onset was randomly jittered across trials, making it impossible for subjects to predict individual-trial repetition onsets, but only the range of onsets. To be ready for repetition detection, subjects needed to select the relevant auditory input just in time, before the first possible repetition, hence the temporal location and specificity of the effect. Since repetitions took place over such a wide temporal range (~2.5 s), the transient attentional selection locked to stimulus onset is inevitably not locked to repetition onset. In other words, the significant alpha decoding in the stimulus-locked analysis reflects anticipation of the repetition onset driven by temporal expectations of the repetition, rather than it being evoked by the actual repetition itself.

The cortical pattern underlying significant alpha classification in the early interval in our repetition-locked analysis showed some resemblance to the pattern observed for the stimulus-locked analysis, with an even more prominent weighting of auditory areas (compare Fig. 3*D* with the left maps in Fig. 4*E*). To investigate whether the significant stimulus- and repetition-locked alpha classification effects (Figs. 3*B*,*C*, 4*B*,*C*) reflect the same or different neurocognitive processes, we trained the classifier on data in the prerepetition interval locked to stimulus onset, and tested it on data in the postrepetition interval locked to repetition onset (i.e., GAT; King and Dehaene, 2014). Importantly, we did not observe any significant across-interval classification (Fig. 5), indicating nonidentical neurocognitive processes before and after repetition onset. Note, however, that the repetition-locked classification results likely reflect a mixture of processes related to attending to the correct stream, detecting the repetition, and responding to the repetition. Based on the nonsignificant across-interval classification illustrated in Figure 5, we cannot fully exclude the possibility that at least part of the mixture of processes after repetition onset overlaps with the process of attending to the cued auditory input before repetition onset.

In addition to in the alpha band, we also observed high classifier performance in the delta band in between repetition onset and the average response (from 0.32 to 1.04 s, from 0.34 to 0.98 s, and from 0.40 to 0.92 s relative to repetition onset; cluster corrected at $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively; Fig. 4*B*, gray line). Paralleling the alpha band, classifier performance was

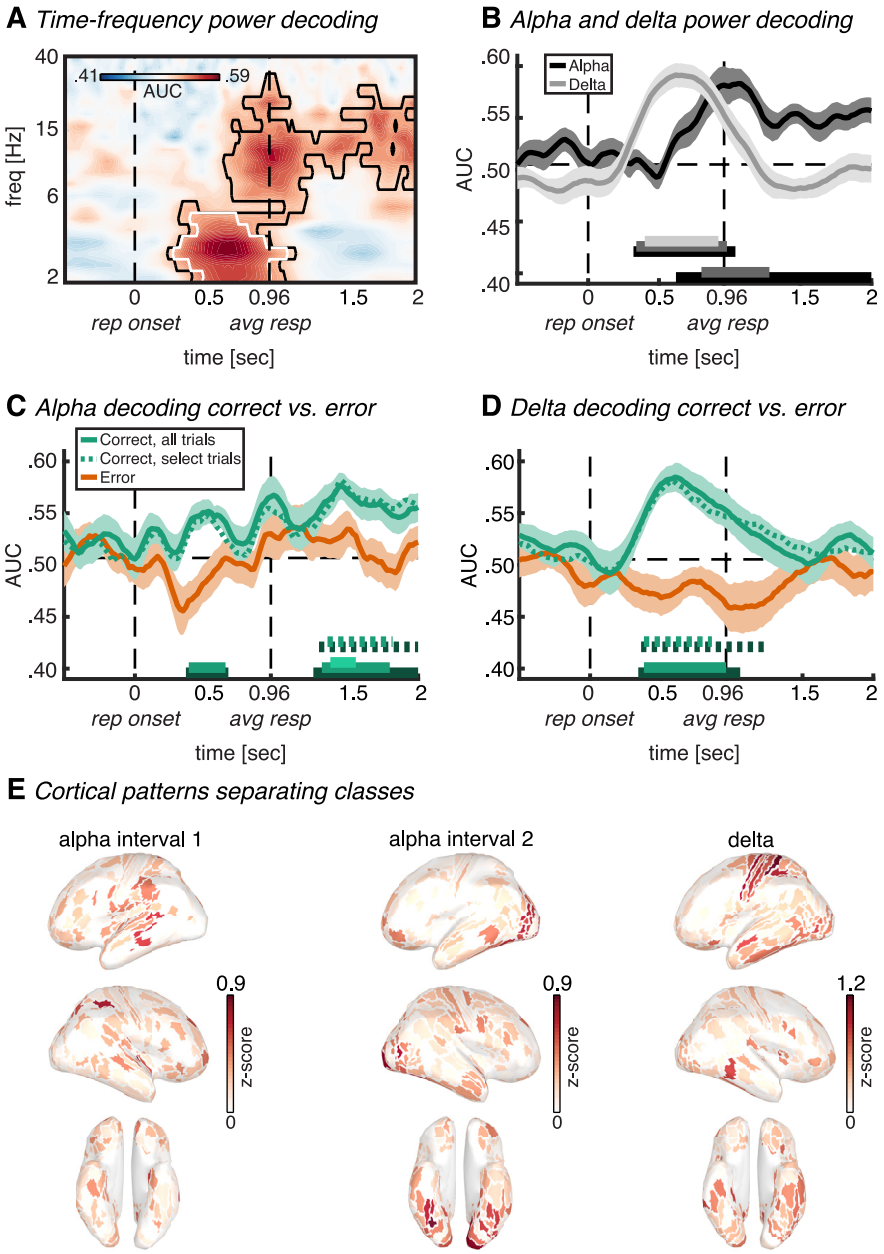

Figure 4. Repetition-locked decoding. ***A***, Time–frequency map of classifier accuracy (AUC). Outlines indicate significant classification at $p < 0.05$ (black) and $p < 0.01$ (white), cluster corrected. ***B***, Classifier accuracy plotted over time for the alpha band (8–14 Hz; black line) and delta band (2–5 Hz; gray line). Horizontal bars on the *x*-axis indicate significant classification after cluster correction for alpha (bottom right duplet) and delta (top left triplet), at $p < 0.05$ (black), $p < 0.01$ (dark gray), or $p < 0.001$ (light gray). ***C***, Alpha power classifier accuracy as in ***B***, but with separate classification analysis for correct (green) or error trials (orange). The dotted green line signifies classification using a random selection of correct trials equal to the number of error trials per individual subject. Green horizontal bars (full and dotted) denote significant cluster-corrected classification (dark green, $p < 0.05$; middle-green, $p < 0.01$; light green, $p < 0.001$). ***D***, Same as ***C***, but now for delta power. Thick lines and shaded areas in ***B***, ***C***, and ***D*** denote subject mean and SEM, respectively. ***E***, Left, middle, Cortical maps of activation patterns separating classes for alpha power in the first and second significant ($p < 0.01$) interval, respectively, as indicated in ***C*** by middle green horizontal bars. Right, Delta power class separation in the significant ($p < 0.01$) interval indicated in ***D*** by the light green horizontal bar. Maps were *z*-scored before averaging, and absolute values were taken to highlight contributing ROIs. Left lateral, right lateral, and ventral views are shown. *rep onset*, Repetition onset; *avg resp*, time of response averaged over subjects and trials.

significant when including only correct trials (from 0.34 to 1.06 s and from 0.38 to 0.96 s relative to repetition onset; cluster corrected at $p < 0.05$ and $p < 0.01$, respectively; Fig. 4*D*, green line), but not when including only error trials (Fig. 4*D*, orange line). Again, a direct comparison between correct and error trials resulted in a significant difference from 0.32 to 1.30 s relative to repetition onset (cluster corrected at $p < 0.05$). One possibility is that since speech entrainment falls in the delta-to-theta range, the delta classification observed here reflects the dissociation between the presence and absence of speech tracking when the speech signal is either cued or uncued, respectively. It is, however, surprising that speech tracking would start only after repetition onset, and not, as our stimulus-locked analysis suggests, in anticipation of the expected repetition onset. An alternative interpretation is that delta classification here reflects initiation and execution of the button press on detection of the repetition. Subjects responded slightly faster on attend speech trials compared with attend environment trials. The classifier may have picked up subtle differences in the neural processes underlying these response time differences. In other words, significant delta classification here might simply reflect the response itself, rather than being related to auditory attention. The cortical patterns of activity underlying significant classification seem to favor the latter explanation. That is, the pattern underlying significant classification from delta (Fig. 4*E*, right) has strongest weighting in motor-related areas. Since the repetition-locked classification results likely reflect a mixture of neurocognitive processes related to attending to the correct stream, detecting the repetition, and responding to the repetition, we did not perform exploratory correlation analyses similar to those for our stimulus-locked analyses between decoding accuracy and behavior, as their meaning would be ambiguous.

Control analyses

To ensure that the significant classification of time-frequency-specific power reported here really reflects oscillatory activity, we performed two sets of control analyses on the main results of stimulus- and repetition-locked alpha and delta power classification. First, we compared our main results (Figs. 3*B*, 4*B*), with classification based on non-wavelet-convolved data. That is, we used the single-trial source-reconstructed neural activity patterns (time domain) as input to the classifier, without applying wavelet convolution. Second, we compared our main results, which were based on total power, with classification based on induced (non-phase-locked) power. Induced power was acquired by subtracting the ERF from each individual trial before performing wavelet convolution (Cohen, 2014).

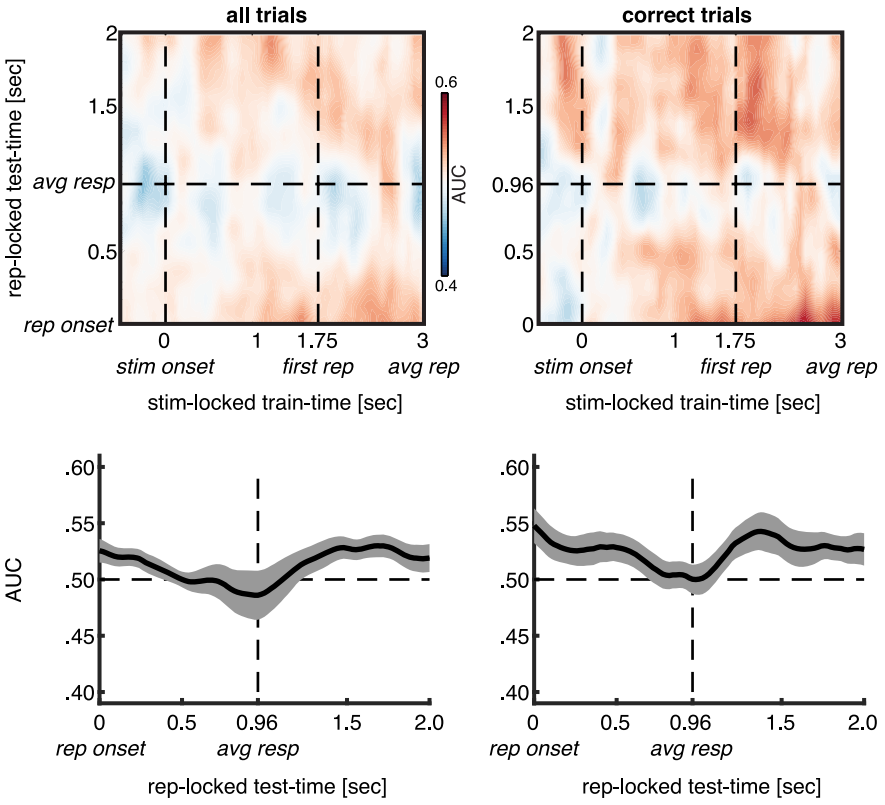

Figure 5. Across-interval decoding. Top row, GAT plots for which the classifier was trained on data in the prerepetition interval locked to stimulus onset and was tested on data in the postrepetition interval locked to repetition onset. Bottom row, Same, but averaged over the stimulus-locked training time within the interval of significant classification at $p < 0.01$ in the main stimulus-locked analysis (Fig. 3*B*,*C*). Thick lines and shaded areas denote subject mean and SEM, respectively. Left and right column are based on all trials or only on trials in which the repetition was detected, respectively. *stim onset*, Auditory stimulus onset; *first rep*, time of first repetition averaged over subjects; *avg rep*, time of repetition averaged over subjects and trials; *rep onset*, repetition onset; *avg resp*, time of response averaged over subjects and trials.

Most importantly, the two sets of control analyses unambiguously confirm that oscillatory activity underlies the significant classification observed here, rather than it being a mere reflection of the stimulus-evoked response. First, classification analyses based on non-wavelet-convolved broadband activity did not show any above-chance classification accuracy in either the stimulus-locked (Fig. 6*A*, left) or the repetition-locked (Fig. 6*B*, left) analyses. Second, classification based on induced alpha power showed a qualitatively similar pattern compared with total alpha power in the stimulus-locked analyses, with a longer significant interval at the $p < 0.05$ threshold, and a significant interval at the $p < 0.001$ threshold, which was not present for total power (significant from 1.32 to 2.04 s, from 1.70 to 1.96 s, and from 1.78 to 1.94 s; cluster corrected at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively; Fig. 6*A*, middle). Similarly, in the repetition-locked analyses, classification accuracy showed a very similar pattern between induced and total alpha power, with an additional extended significant interval at the $p < 0.01$ threshold (significant from 0.74 to 2.00 s at $p < 0.05$, and from 0.82 to 1.24 and 1.34 to 2.00 s; cluster corrected at $p < 0.01$; Fig. 6*B*, middle). Last, classification accuracy based on delta power was also comparable between induced and total power (from 0.34 to 1.02 s, from 0.36 to 0.94 s, and from 0.42 to 0.88 s; cluster corrected at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively; Fig. 6*B*, right). The fact that induced power resulted in the same pattern as total power indicates that the effect in alpha observed here reflects a modulation in a brain-inherent rhythm that is not phase locked to the auditory stimulation.

Wavelet convolution inevitably results in some temporal smearing of the resulting power estimation. The concern here with temporal smearing would be that our peak in alpha classification in the stimulus-locked analysis does not actually reflect temporal anticipation of repetition-onset, but rather is induced by the repetition and temporally smear back in the time to before repetition onset. However, in that case one would expect to observe significant alpha classification postrepetition in the repetition-locked analysis, which we did not. Furthermore, note that, given the Gaussian shape of a wavelet, the weighting is strongest at its center, and it is only to a limited extent influenced by surrounding time points. Nevertheless, we performed a control analysis taking the temporal smearing because of wavelet convolution into account (Fig. 6*C*). Most importantly, the patterns look qualitatively very similar, and, although it is based on fewer trials, classification accuracy is still significant.

To test for a potential confound of a difference in task difficulty between attend speech and attend environment trials on classification accuracy, we performed two control analyses on our main results of significant stimulus-locked alpha classification (Fig. 3*B*,*C*), for which we operationalized task difficulty as behavioral accuracy (i.e., detection rate). First, we performed the exact same analysis as our main analysis using all trials (Fig. 3*B*), but now after equalizing the ratio of correct trials to error trials between the two classes (Fig. 6*D*). Alpha classification was still significant from 1.66 to 2.04 s relative to stimulus onset (cluster corrected at $p < 0.05$), which indicates that performance was not a confounding factor. Note that our analysis using only correct trials (Fig. 3*C*, green line) is inherently balanced concerning performance, and thus additionally confirms that performance was not a confounding factor. Second, we correlated across subjects the performance difference between attend speech and attend environment trials with the classification accuracy averaged over the interval of significant alpha classification ($p < 0.01$; Fig. 3*B*, gray horizontal bar). We did not observe any significant correlation for behavioral accuracy (Rho = 0.22, $p = 0.43$) or for response time (Rho = −0.05, $p = 0.85$). These correlation coefficients indicate that the difference in behavioral performance can explain only a small portion of the variance in classification accuracy, suggesting that a difference in task difficulty between stimuli is not the main driver of the current results. In contrast, previous studies suggesting sensitivity of univariate alpha power modulations to task difficulty or subjective effort did find such a correlation between the two measures (Obleser et al., 2012; Wöstmann et al., 2015). Together, while we cannot fully exclude the possibility that task difficulty or subjective effort played a minor role, the control analyses performed here give converging support that significant classification was not driven by it.

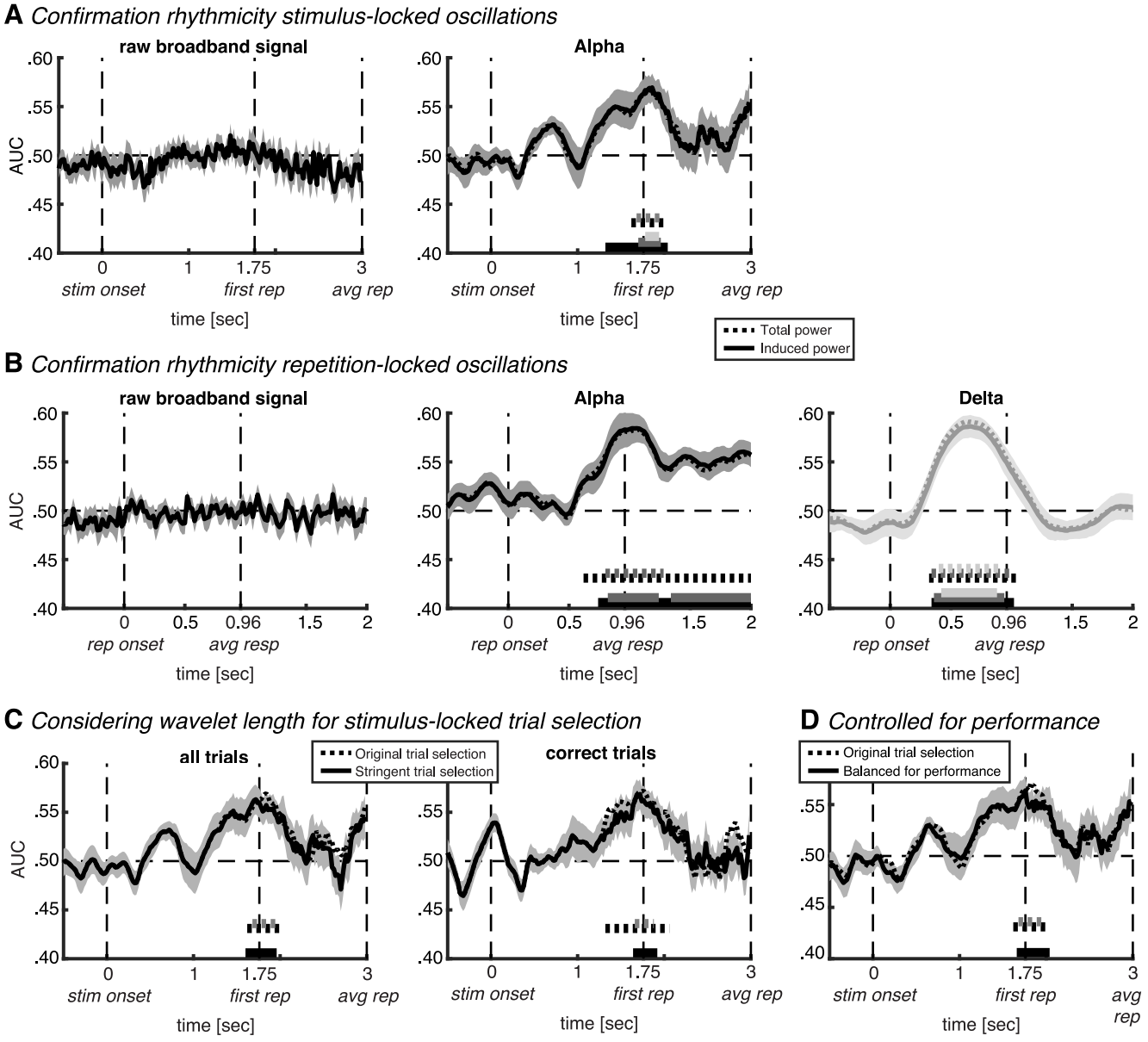

Figure 6. Control analyses. ***A***, Confirmation of rhythmicity of alpha oscillations in stimulus-locked analysis. Left, Classifier accuracy plotted over time for non-wavelet-convolved neural activity. Right, Comparison between our original result based on total alpha power (dotted line) and induced (i.e., non-phase locked) alpha power (full line). ***B***, Same control analyses as in ***A***, but repetition locked, and in both alpha power (black lines, middle) and delta power (light gray lines, right). ***C***, More stringent trial rejection at individual-trial repetition onset, taking the wavelet length into account (see main text for details). ***D***, Control for potential confounding factor of task difficulty (operationalized as behavioral accuracy), by taking into account between-class differences in performance (see main text for details). In all plots, thick lines and shaded areas denote subject mean and SEM, respectively. Horizontal (full and dotted) bars on the *x*-axis indicate significant classification after cluster correction at $p < 0.05$ (black), $p < 0.01$ (dark gray), or $p < 0.001$ (light gray). *stim onset*, Auditory stimulus onset; *first rep*, time of first repetition averaged over subjects; *avg rep*, time of repetition averaged over subjects and trials; *rep onset*, repetition onset; *avg resp*, time of response averaged over subjects and trials.

Discussion

We investigated the neural mechanisms of object-based auditory attention. We recorded MEG while subjects prepared for repetition of an auditory object presented in one of two spatially and temporally overlapping naturalistic auditory streams (i.e., speech and environmental). Subjects were cued with 70% validity in which stream the repetition would appear and were instructed to press a button on detection. We trained a linear classifier on the cortical distribution of source-reconstructed oscillatory activity to distinguish which auditory stream was attended in anticipation of the repetition. First, we could classify which auditory stream was attended from alpha oscillations in anticipation of repetition onset. These anticipatory alpha oscillations played an important role in successful detection of the repetition, as indicated by significant classification from trials in which subjects subsequently detected the repetition, but not from miss trials. This was further supported by a negative correlation between classification accuracy and reaction time: subjects showing higher classification accuracy right before repetition onset were faster in responding to the repetition. Second, in a repetition-locked analysis we similarly observed significant classification from alpha in between repetition onset and response for correct trials, but not for error trials. Last, we observed classification from delta power after repetition onset, again for correct trials only, with a cortical distribution that underscored motor-related

areas. These delta oscillations presumably reflected the planning, initiation, and/or execution of the button press.

Object-based auditory attention

The literature on auditory attention has largely focused on improving the performance of auditory attention decoders (Miran et al., 2018; Wong et al., 2018; Alickovic et al., 2019; Etard et al., 2019; Taillez et al., 2020), often for real-world applications such as hearing aids (O'Sullivan et al., 2017; Han et al., 2019). While invaluable, these studies do not delineate the underlying top-down control mechanisms or unveil which components of the neural signal contribute to successful classification. Most paradigms that investigate top-down control of auditory attention use highly simplified and controlled stimuli, usually allowing attention to gear toward a specific manipulated feature. For example, in the standard dichotic listening paradigm a lateralized cue indicates the relevance of individual streams presented to each ear, making it easy to distinguish the two based on spatial information (Ahveninen et al., 2013). Other successful paradigms manipulated features such as pitch (Hill and Miller, 2010) or background noise (Ding and Simon, 2013). However, naturalistic auditory scenes we typically encounter in daily life are a complex mixture of spatially and temporally overlapping sounds, making it hard to separate relevant from irrelevant based on a single feature. Instead, attention has to operate on auditory objects of interest (Griffiths and Warren, 2004), much akin to object-based attention, as described in the visual attention literature (Roelfsema et al., 1998; Baldauf and Desimone, 2014).

Crucially, the current auditory repetition detection task necessitated processing the acoustic streams to a cognitive level that allowed for the recognition of a temporally extended set of low-level features as an object and to understand that this object was repeated (Marinato and Baldauf, 2019). To capture the neural dynamics of anticipatory object-based auditory attention, we adopted an MVPA approach. The tuning of neuronal excitability for target facilitation and distractor suppression occurs simultaneously in neighboring subregions of higher auditory cortex, thus resulting in a complex and distributed activity pattern. For example, alpha suppression in cortical regions that process relevant auditory input facilitates processing thereof (Leske et al., 2015; Griffiths et al., 2019), while alpha enhancement attenuates processing of distracting stimuli (Strauß et al., 2014; Wöstmann et al., 2017). While careful independent manipulation of the spatial or temporal characteristics of targets and distractors verifies both these processes (Wöstmann et al., 2019a; Deng et al., 2020), this does not reflect a naturalistic auditory scene. However, since neural signals reflecting these various processes might cancel each other out using traditional univariate methods, it is difficult to disentangle facilitating and inhibiting effects with the spatial resolution of MEG. MVPA instead is sensitive to such subtle differences in brain states (Stokes et al., 2015; de Vries et al., 2019). Importantly, with the current approach we were indeed able to classify object-based auditory attention from oscillatory MEG activity in the alpha band, in anticipation of the expected auditory object of interest. The observed cortical distribution of activation patterns verified a complex contribution of multiple higher-level auditory areas (Puvvada and Simon, 2017; O'Sullivan et al., 2019); several visual areas that have been shown to activate during auditory attention (Cate et al., 2009; He et al., 2013; Vetter et al., 2014; Alho et al., 2015; Bi et al., 2016), such as areas involved in multimodal scene perception (e.g., PPA; Bedny, 2017); and frontal executive-control areas involved in scene navigation (Vann et al., 2009) or the monitoring of expected events (Petrides, 2005).

Alpha oscillations in auditory attention

We observed here significant anticipatory attention classification specifically in the alpha frequency band. Ample work has shown the importance of prestimulus alpha oscillations in sensory areas for the top-down control of visual attention (Klimesch et al., 2007; Bagherzadeh et al., 2020), tactile attention (Haegens et al., 2011a), the prioritization of information within working memory (de Vries et al., 2017, 2018, 2020; Weisz et al., 2020), and, important here, auditory attention (Weisz and Obleser, 2014). In fact, evidence suggests that alpha modulations directly affect neural processing of auditory input. For instance, alpha modulations in auditory cortex predict the subsequent stimulus-evoked response (Wöstmann et al., 2019b), attentional gain of cortical representations of attended speech (Kerlin et al., 2010), and speech tracking by delta-to-theta oscillations (Keitel et al., 2017; but see Hauswald et al., 2020). The fact that here classification was significant only if subjects subsequently detected the repetition, and that classification accuracy correlated with performance on the speeded detection task, further testifies to the functional relevance of anticipatory alpha oscillations for object-based auditory attention. Similarly, prestimulus alpha modulations predict subsequent near-threshold auditory perception (Leske et al., 2015; Herrmann et al., 2016), spatial pitch discrimination (Wöstmann et al., 2019a), confidence on an auditory discrimination task (Wöstmann et al., 2019b), and speech intelligibility (Obleser and Weisz, 2012; Hauswald et al., 2020). Albeit speculative, these and our current results suggest a causal role for alpha oscillations in auditory attention.

We add to this literature by showing that alpha oscillations are not only important for attention to specific low-level features in simple stimuli, but also for successful object-based auditory attention in complex naturalistic auditory scenes. Given that alpha oscillations are believed to index the excitability of underlying neuronal populations in sensory areas (Haegens et al., 2011b, 2015), we propose that the classification of anticipatory attention from alpha oscillations observed here reflects a tuning of the neuronal excitability in the relevant cortical subregions involved in encoding and processing the to-be-attended stream, and inhibiting the to-be-ignored stream. This also demonstrates that the spatial scale at which the speech and environmental signals used here are processed is large enough to be picked up with multivariate source-reconstructed MEG analyses. Interestingly, classification was not sustained throughout stimulus presentation or throughout the temporal range of possible repetition onsets. Rather, it peaked right before the first expected repetition onset, in line with findings on temporal attention (Nobre and Van Ede, 2018; Wöstmann et al., 2021), which indicates that object-based auditory attention acts as a transient selection mechanism according to temporal expectations of its necessity, rather than a continuous tracking mechanism.

Caveats

Note that alpha oscillations presumably do not carry the auditory content itself, which is likely carried by delta, theta, gamma, and theta–gamma coupling (Lakatos et al., 2005; Giraud and Poeppel, 2012). Rather, alpha oscillations reflect top-down attentional control (Hartmann et al., 2014; Bagherzadeh et al., 2020; Weisz et al., 2020), which modulates the neuronal excitability of regions that will process the upcoming auditory information. A

potential general limitation of time/frequency-resolved MEG/EEG signals is that oscillations might reflect the time-resolved stimulus-evoked response (i.e., the ERF), rather than true oscillatory activity. However, in our control analyses we show that it was not possible to decode auditory attention from the time-resolved signal, and that classification accuracy did not suffer from removing the ERF from each single trial before time–frequency analysis (i.e., induced power; Cohen, 2014). The results of these two control analyses give further support for the specific importance of oscillatory brain activity in the alpha frequency range for object-based auditory attention. Last, note that while the repetition-locked analysis shows a significant cluster-corrected difference in alpha classification between correct and error trials, in the stimulus-locked analysis we observed only an uncorrected difference, which should therefore be interpreted with care.

Conclusions

To conclude, our results reveal that a complex distributed cortical pattern of alpha oscillations underlies successful object-based auditory top-down attention and indicate that alpha oscillations operate in a transient and timely manner depending on temporal expectations about an anticipated auditory object of interest. On a more general level, these results are a testament to the use of multivariate analysis methods on time–frequency-decomposed, source-reconstructed MEG/EEG data for investigating higher-level human cognition in general, and object-based attention in particular.

References

Ahveninen J, Huang S, Belliveau JW, Chang W-T, Hämäläinen M (2013) Dynamic oscillatory processes governing cued orienting and allocation of auditory attention. J Cogn Neurosci 25:1926–1943.

Alavash M, Tune S, Obleser J (2019) Modular reconfiguration of an auditory control brain network supports adaptive listening behavior. Proc Natl Acad Sci U S A 116:660–669.

Alho K, Salmi J, Koistinen S, Salonen O, Rinne T (2015) Top-down controlled and bottom-up triggered orienting of auditory attention to pitch activate overlapping brain networks. Brain Res 1626:136–145.

Alickovic E, Lunner T, Gustafsson F, Ljung L (2019) A tutorial on auditory attention identification methods. Front Neurosci 13:153.

Bagherzadeh Y, Baldauf D, Pantazis D, Desimone R (2020) Alpha synchrony and the neurofeedback control of spatial attention. Neuron 105:577–587.e5.

Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. Science 344:424–427.

Bamiou DE, Musiek FE, Luxon LM (2003) The insula (Island of Reil) and its role in auditory processing: literature review. Brain Res Brain Res Rev 42:143–154.

Banerjee S, Snyder AC, Molholm S, Foxe JJ (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.

Bedny M (2017) Evidence from blindness for a cognitively pluripotent cortex. Trends Cogn Sci 21:637–648.

Bi Y, Wang X, Caramazza A (2016) Object domain and modality in the ventral visual pathway. Trends Cogn Sci 20:282–290.

Cate AD, Herron TJ, Yund EW, Stecker GC, Rinne T, Kang X, Petkov CI, Disbrow EA, Woods DL (2009) Auditory attention activates peripheral visual cortex. PLoS One 4:e4645.

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

Cohen MX (2014) Analyzing neural time series data: theory and practice. Cambridge, MA: MIT.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I: segmentation and surface reconstruction. Neuroimage 9:179–194.

de Vries IEJ, van Driel J, Olivers CNL (2017) Posterior α EEG dynamics dissociate current from future goals in working memory guided visual search. J Neurosci 37:1591–1603.

de Vries IEJ, van Driel J, Karacaoglu M, Olivers CNL (2018) Priority switches in visual working memory are supported by frontal delta and posterior alpha interactions. Cereb Cortex 28:4090–4104.

de Vries IEJ, van Driel J, Olivers CNL (2019) Decoding the status of working memory representations in preparation of visual selection. Neuroimage 191:549–559.

de Vries IEJ, Slagter HA, Olivers CNL (2020) Oscillatory control over representational states in working memory. Trends Cogn Sci 24:150–162.

Deng Y, Choi I, Shinn-Cunningham B (2020) Topographic specificity of alpha power during auditory spatial attention. Neuroimage 207:116360.

Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A 109:11854–11859.

Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33:5728–5735.

Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8:311.

Etard O, Kegler M, Braiman C, Forte AE, Reichenbach T (2019) Decoding of selective attention to continuous speech from the human auditory brainstem response. Neuroimage 200:1–11.

Fahrenfort JJ, Grubert A, Olivers CNL, Eimer M (2017) Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. Sci Rep 7:1886.

Fahrenfort JJ, van Driel J, van Gaal S, Olivers CNL (2018) From ERPs to MVPA using the Amsterdam Decoding and Modeling toolbox (ADAM). Front Neurosci 12:368.

Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195–207.

Frey JN, Mainy N, Lachaux J-P, Müller N, Bertrand O, Weisz N (2014) Selective modulation of auditory cortical alpha activity in an audiovisual spatial attention task. J Neurosci 34:6634–6639.

Fuentemilla L, Penny WD, Cashdollar N, Bunzeck N, Düzel E (2010) Theta-coupled periodic replay in working memory. Curr Biol 20:606–612.

Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511–517.

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC (2016) A multi-modal parcellation of human cerebral cortex. Nature 536:171–178.

Griffiths BJ, Mayhew SD, Mullinger KJ, Jorge J, Charest I, Wimber M, Hanslmayr S (2019) Alpha/beta power decreases track the fidelity of stimulus specific information. Elife 8:e49562.

Griffiths TD, Warren JD (2004) What is an auditory object? Nat Rev Neurosci 5:887–892.

Haegens S, Zion Golumbic E (2018) Rhythmic facilitation of sensory processing: a critical review. Neurosci Biobehav Rev 86:150–165.

Haegens S, Händel BF, Jensen O (2011a) Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. J Neurosci 31:5197–5204.

Haegens S, Nácher V, Luna R, Romo R, Jensen O (2011b) α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. Proc Natl Acad Sci U S A 108:19377–19382.

Haegens S, Cousijn H, Wallis G, Harrison PJ, Nobre AC (2014) Inter- and intra-individual variability in alpha peak frequency. Neuroimage 92:46–55.

Haegens S, Barczak A, Musacchia G, Lipton ML, Mehta AD, Lakatos P, Schroeder CE (2015) Laminar profile and physiology of the α rhythm in primary visual, auditory, and somatosensory regions of neocortex. J Neurosci 35:14341–14352.

Hämäläinen MS, Ilmoniemi RJ (1994) Interpreting magnetic fields of the brain: minimum norm estimates. Med Biol Eng Comput 32:35–42.

Han C, O'Sullivan J, Luo Y, Herrero J, Mehta AD, Mesgarani N (2019) Speaker-independent auditory attention decoding without access to clean speech sources. Sci Adv 5:eaav6134.

Hand DJ, Till RJ (2001) A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach Learn 45:171–186.
Hartmann T, Lorenz I, Müller N, Langguth B, Weisz N (2014) The effects of neurofeedback on oscillatory processes related to tinnitus. Brain Topogr 27:149–157.
Haufe S, Meinecke F, Görgen K, Dähne S, Haynes JD, Blankertz B, Bießmann F (2014) On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87:96–110.
Hauswald A, Keitel A, Chen YP, Rösch S, Weisz N (2020) Degradation levels of continuous speech affect neural speech tracking and alpha power differently. Eur J Neurosci.
He C, Peelen MV, Han Z, Lin N, Caramazza A, Bi Y (2013) Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. Neuroimage 79:1–9.
Herrmann B, Henry MJ, Haegens S, Obleser J (2016) Temporal expectations and neural amplitude fluctuations in auditory cortex interactively influence perception. Neuroimage 124:487–497.
Hill KT, Miller LM (2010) Auditory attentional control and selection during cocktail party listening. Cereb Cortex 20:583–590.
Jafarpour A, Horner AJ, Fuentemilla L, Penny WD, Duzel E (2013) Decoding oscillatory representations and mechanisms in memory. Neuropsychologia 51:772–780.
Keitel A, Ince RAA, Gross J, Kayser C (2017) Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. Neuroimage 147:32–42.
Kerlin JR, Shahin AJ, Miller LM (2010) Attentional gain control of ongoing cortical speech representations in a "cocktail party". J Neurosci 30:620–628.
King JR, Dehaene S (2014) Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn Sci 18:203–210.
Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res Rev 53:63–88.
Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904–1911.
Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal filter mechanism of auditory selective attention. Neuron 77:750–761.
Lee AKC, Rajaram S, Xia J, Bharadwaj H, Larson E, Hämäläinen MS, Shinn-Cunningham BG (2013) Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. Front Neurosci 6:190.
Lee T-W, Girolami M, Sejnowski TJ (1999) Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Comput 11:417–441.
Leske S, Ruhnau P, Frey J, Lithari C, Müller N, Hartmann T, Weisz N (2015) Prestimulus network integration of auditory cortex predisposes near-threshold perception independently of local excitability. Cereb Cortex 25:4898–4907.
Marinato G, Baldauf D (2019) Object-based attention in complex, naturalistic auditory streams. Sci Rep 9:2854.
Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.
Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485:233–236.
Miran S, Akram S, Sheikhattar A, Simon JZ, Zhang T, Babadi B (2018) Real-time tracking of selective auditory attention from M/EEG: a Bayesian filtering approach. Front Neurosci 12:262.
Müller N, Weisz N (2012) Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. Cereb Cortex 22:1604–1613.
Müller N, Lorenz I, Langguth B, Weisz N (2013) rTMS induced tinnitus relief is related to an increase in auditory cortical alpha activity. PLoS One 8: e55557.
Newman EL, Norman KA (2010) Moderate excitation leads to weakening of perceptual representations. Cereb Cortex 20:2760–2770.
Nobre AC, Van Ede F (2018) Anticipated moments: temporal structure in attention. Nat Rev Neurosci 19:34–48.
Obleser J, Weisz N (2012) Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cereb Cortex 22:2466–2477.
Obleser J, Wöstmann M, Hellbernd N, Wilsch A, Maess B (2012) Adverse listening conditions and memory load drive a common alpha oscillatory network. J Neurosci 32:12376–12383.
Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.
O'Sullivan J, Chen Z, Herrero J, McKhann GM, Sheth SA, Mehta AD, Mesgarani N (2017) Neural decoding of attentional selection in multi-speaker environments without access to clean sources. J Neural Eng 14:56001.
O'Sullivan J, Herrero J, Smith E, Schevon C, McKhann GM, Sheth SA, Mehta AD, Mesgarani N (2019) Hierarchical encoding of attended auditory objects in multi-talker speech perception. Neuron 104:1195–1209.e3.
Petrides M (2005) Lateral prefrontal cortex: architectonic and functional organization. Philos Trans R Soc Lond B Biol Sci 360:781–795.
Praamstra P, Kourtis D, Hoi FK, Oostenveld R (2006) Neurophysiology of implicit timing in serial choice reaction-time performance. J Neurosci 26:5448–5455.
Puvvada KC, Simon JZ (2017) Cortical representations of speech in a multi-talker auditory scene. J Neurosci 37:9189–9196.
Ratcliff R (1993) Methods for dealing with reaction time outliers. Psychol Bull 114:510–532.
Roelfsema PR, Lamme VAF, Spekreijse H (1998) Object-based attention in the primary visual cortex of the macaque monkey. Nature 395:376–381.
Rohenkohl G, Nobre AC (2011) Alpha oscillations related to anticipatory attention follow temporal expectations. J Neurosci 31:14076–14084.
Sauseng P, Klimesch W, Stadler W, Schabus M, Doppelmayr M, Hanslmayr S, Gruber WR, Birbaumer N (2005) A shift of visual spatial attention is selectively associated with human EEG alpha activity. Eur J Neurosci 22:2917–2926.
Shinn-Cunningham BG (2008) Object-based auditory and visual attention. Trends Cogn Sci 12:182–186.
Stokes MG, Wolff MJ, Spaak E (2015) Decoding rich spatial information with high temporal resolution. Trends Cogn Sci 19:636–638.
Strauß A, Wöstmann M, Obleser J (2014) Cortical alpha oscillations as a tool for auditory selective inhibition. Front Hum Neurosci 8:350.
Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM (2011) Brainstorm: a user-friendly application for MEG/EEG analysis. Comput Intell Neurosci 2011:879716.
Tadel F, Bock E, Niso G, Mosher JC, Cousineau M, Pantazis D, Leahy RM, Baillet S (2019) MEG/EEG group analysis with Brainstorm. Front Neurosci 13:76.
Taillez T, Kollmeier B, Meyer BT (2020) Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech. Eur J Neurosci 51:1234–1241.
Taulu S, Simola J (2006) Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys Med Biol 51:1759.
Thut G, Nietzel A, Brandt SA, Pascual-Leone A (2006) α-Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. J Neurosci 26:9494–9502.
Vaden KI, Kuchinsky SE, Cute SL, Ahlstrom JB, Dubno JR, Eckert MA (2013) The cingulo-opercular network provides word-recognition benefit. J Neurosci 33:18979–18986.
van Driel J, Ort E, Fahrenfort JJ, Olivers CNL (2019) Beta and theta oscillations differentially support free versus forced control over multiple-target search. J Neurosci 39:1733–1743.
Van Ede F, De Lange F, Jensen O, Maris E (2011) Orienting attention to an upcoming tactile event involves a spatially and temporally specific modulation of sensorimotor alpha- and beta-band oscillations. J Neurosci 31:2016–2024.
van Moorselaar D, Theeuwes J, Olivers CN (2014) In competition for the attentional template: can multiple items within visual working memory guide attention? J Exp Psychol Hum Percept Perform 40:1450–1464.
Vann SD, Aggleton JP, Maguire EA (2009) What does the retrosplenial cortex do? Nat Rev Neurosci 10:792–802.
Vetter P, Smith FW, Muckli L (2014) Decoding sound and imagery content in early visual cortex. Curr Biol 24:1256–1262.
Weisz N, Obleser J (2014) Synchronisation signatures in the listening brain: a perspective from non-invasive neuroelectrophysiology. Hear Res 307:16–28.

Weisz N, Hartmann T, Müller N, Lorenz I, Obleser J (2011) Alpha rhythms in audition: cognitive and clinical perspectives. Front Psychol 2:73.

Weisz N, Kraft NG, Demarchi G (2020) Auditory cortical alpha/beta desynchronization prioritizes the representation of memory items during a retention period. Elife 9:e55508.

Wong DDE, Fuglsang SA, Hjortkjær J, Ceolini E, Slaney M, de Cheveigné A (2018) A comparison of regularization methods in forward and backward models for auditory attention decoding. Front Neurosci 12:531.

Wöstmann M, Herrmann B, Wilsch A, Obleser J (2015) Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. J Neurosci 35:1458–1467.

Wöstmann M, Lim SJ, Obleser J (2017) The human neural alpha response to speech is a proxy of attentional control. Cereb Cortex 27:3307–3317.

Wöstmann M, Vosskuhl J, Obleser J, Herrmann CS (2018) Opposite effects of lateralised transcranial alpha versus gamma stimulation on auditory spatial attention. Brain Stimul 11:752–758.

Wöstmann M, Alavash M, Obleser J (2019a) Alpha oscillations in the human brain implement distractor suppression independent of target selection. J Neurosci 39:9797–9805.

Wöstmann M, Waschke L, Obleser J (2019b) Prestimulus neural alpha power predicts confidence in discriminating identical auditory stimuli. Eur J Neurosci 49:94–105.

Wöstmann M, Maess B, Obleser J (2021) Orienting auditory attention in time: lateralized alpha power reflects spatio-temporal filtering. Neuroimage 228:117711.

Zanto TP, Pan P, Liu H, Bollinger J, Nobre AC, Gazzaley A (2011) Age-related changes in orienting attention in time. J Neurosci 31:12461–12470.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". Neuron 77:980–991.