Neural Dynamics Underlying Attentional Orienting to Auditory Representations in Short-Term Memory

Kristina C. Backer,^{1,2} Malcolm A. Binns,^{1,3} and Claude Alain^{1,2}

¹Rotman Research Institute at Baycrest Centre, Toronto, Ontario, M6A 2E1, Canada, and ²Department of Psychology and ³Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, M5S 1A1, Canada

Sounds are ephemeral. Thus, coherent auditory perception depends on “hearing” back in time: retrospectively attending that which was lost externally but preserved in short-term memory (STM). Current theories of auditory attention assume that sound features are integrated into a perceptual object, that multiple objects can coexist in STM, and that attention can be deployed to an object in STM. Recording electroencephalography from humans, we tested these assumptions, elucidating feature-general and feature-specific neural correlates of auditory attention to STM. Alpha/beta oscillations and frontal and posterior event-related potentials indexed feature-general top-down attentional control to one of several coexisting auditory representations in STM. Particularly, task performance during attentional orienting was correlated with alpha/low-beta desynchronization (i.e., power suppression). However, attention to one feature could occur without simultaneous processing of the second feature of the representation. Therefore, auditory attention to memory relies on both feature-specific and feature-general neural dynamics.

Key words: alpha/beta oscillations; attention; auditory; EEG; feature; object

Introduction

Sounds unfold and disappear across time. Consequently, listeners must maintain and attend to memory representations of what was heard, enabling coherent auditory perception. This suggests that attention to auditory representations in short-term memory (STM), a form of reflective attention (Chun and Johnson, 2011), is particularly inherent in audition. This idea falls in line with a predominant model, the object-based account of auditory attention (Alain and Arnott, 2000; Shinn-Cunningham, 2008), that makes several predictions regarding how sounds are represented and attended. (1) Sound features are integrated into a perceptual object, whereby attentional deployment to one feature enhances the processing of all features comprising that object (Zatorre et al., 1999; Dyson and Ishfaq, 2008). (2) Multiple sound objects can coexist in STM (Alain et al., 2002). (3) Top-down attention can be allocated to one of several coexisting sound objects in STM (Backer and Alain, 2012).

Studies that used prestimulus cues to direct attention to one of multiple concurrent sound streams provided evidence supporting this model (Kerlin et al., 2010; Ding and Simon, 2012). How-

ever, the predictions of the object-based account may not persist when attention acts within STM, after the initial encoding of concurrent sounds has occurred. For instance, sounds from various physical sources may be encoded and perceived as a Gestalt representation. Alternatively, each sound source may be encoded separately and accessed by a defining feature, such as its identity or location. Auditory semantic and spatial features are thought to be processed by parallel dorsal (“where”) and ventral (“what”) processing pathways (Rauschecker and Tian, 2000; Alain et al., 2001), and attention to a semantic or spatial feature of a sound can modulate the underlying neural networks according to this dual-pathway model (Hill and Miller, 2010). However, it remains unclear whether reflective attention to a spatial or semantic feature affects neural activity in a similar manner.

In this study, we used electroencephalography (EEG) to examine the neural activity underlying the interactions between auditory attention and STM, providing a glimpse into how concurrent sounds are represented in STM and testing the object-based model of auditory attention in a unique way. To this end, we developed an auditory feature–conjunction change detection task with retro-cues (Fig. 1). Retro-cues endogenously directed attention to one of several coexisting auditory representations in STM based on its semantic or spatial feature; uninformative cues instructed the participant to maintain the identity and location of each sound. By comparing EEG dynamics between Informative–Semantic and Informative–Spatial versus Uninformative retro-cue trials, we could isolate domain-general (i.e., feature-general) activity, reflecting the control of attentional orienting to a specific sound object representation in the midst of competing representations. We could also identify domain-specific (i.e., feature-specific) activity by comparing activity during Informative–Semantic versus Informative–Spatial retro-cue trials. We show that both domain-

Received April 11, 2014; revised Oct. 29, 2014; accepted Nov. 4, 2014.

Author contributions: K.C.B. and C.A. designed research; K.C.B. performed research; K.C.B., M.A.B., and C.A. analyzed data; K.C.B., M.A.B., and C.A. wrote the paper.

This work was supported by funding from the Natural Sciences and Engineering Research Council of Canada (a Discovery Grant to C.A. and a CREATE-Auditory Cognitive Neuroscience graduate award to K.C.B.). We thank Yu He and Dean Carcone for assistance with data collection.

The authors declare no competing financial interests.

Correspondence should be addressed to Kristina C. Backer at her present address: Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd Street, Seattle, WA 98105-6246. E-mail: kbacker@uw.edu.

DOI:10.1523/JNEUROSCI.1487-14.2015

Copyright © 2015 the authors 0270-6474/15/351307-12\$15.00/0

Figure 1. This figure depicts examples of an Informative–Semantic retro-cue trial (top) and an Informative–Spatial retro-cue trial (bottom). An example of an Uninformative retro-cue trial is not pictured. Semantic retro-cues took the form of a letter corresponding to a sound category (e.g., “H” referring to the human sound), while spatial retro-cues were presented as a number (e.g., “3”) corresponding to one of the four sound locations. Mem., memory; R-C, retro-cue; ITl, intertrial interval.

general and domain-specific activity contributes to attentional orienting within STM, providing evidence against a strict formulation of the object-based model—specifically, that attention to one feature may not enhance processing of the other feature(s) of this object.

Materials and Methods

The Research Ethics Board at Baycrest Centre approved the study protocol. All participants gave written, informed consent before the commencement of the experiment and received monetary compensation for their time.

Participants

We recruited 18 young adults for the study. At the beginning of the experiment, each volunteer completed an audiogram (octave intervals from 250 to 8000 Hz) to ensure that he/she had clinically normal pure-tone hearing thresholds (i.e., ≤ 25 dB HL at all frequencies in both ears). Data from two participants were excluded; one failed to meet the hearing thresholds criteria and another reported diagnosis of an attention disorder during the session. Thus, the analyses included data from 16 participants (nine females; mean age, 22.2 years; range, 18–31 years).

Stimuli

Everyday sounds were chosen from an in-house database that comprises non-speech human (e.g., laugh, cry, cough), animal (e.g., dog bark, bird chirp, sheep baa), music (e.g., piano tone, guitar strum, drum roll), and man-made object (e.g., siren, electric drill, car honk) sounds. From each of the four categories, we chose 48 exemplars. All sounds were sampled at 12,207 Hz, lasted for 1005 ms, and were normalized for loudness (same root mean square power).

From these sounds, we generated 288 auditory scenes, each comprising four simultaneous sounds (one per category). These scenes were carefully created, taking the spectrotemporal structure of each sound into consideration, with the constraint that no individual sound could be used more than seven times. During a control experiment, which involved eight participants who did not participate in the EEG study, the

288 scenes were played one at a time, followed by a silent delay (~ 1500 ms) and then a spatial retro-cue (a visually presented number corresponding to one of the four sound locations). Participants pressed a button corresponding to the category of the sound that played in the retro-cued location; the 96 “best” scenes (i.e., those to which most participants responded accurately) were chosen for the EEG study.

Each sound within the auditory scenes was played from a different free-field location (-90° , -30° , $+30^\circ$, $+90^\circ$) using JBL speakers positioned ~ 1 m from the participant’s head. Black curtains prevented participants from seeing the speakers during the study. The intensity of the stimuli [on average ~ 60 dB (A-weighted) sound pressure level (SPL)] was measured using a Larson-Davis SPL meter (model 824) with a free-field microphone (Larson-Davis model 2559) placed at the same position where each participant would be sitting.

Procedure and experimental design

First, each participant completed a sound identification task (lasting ~ 15 min) to ensure that he/she could successfully categorize each individual sound played in the experiment. All 180 individual sounds were played, one at a time, from a simulated location at 0° relative to the participant’s line of sight, using the -30° and $+30^\circ$ speakers. Participants pressed a button to indicate their categorization of each sound. Feedback was visually displayed on each trial. We instructed participants to respond as accurately as possible and to make a mental note of the correct answer to any stimuli they labeled incorrectly because they would be categorizing sounds in the main experiment.

After the identification task, participants were instructed on the primary experimental (i.e., change detection) task. The trial structure is illustrated in Figure 1. On each trial, participants heard an auditory scene (1005 ms), followed by a silent delay [interstimulus interval (ISI), 1100–1650 ms], a retro-cue (presented for 500 ms), a second silent delay (ISI, 1450–2000 ms), and finally an auditory probe. Participants pressed a button to indicate whether or not the probe matched its original location (in other words, if the probe matched the sound that originally played at that location), with the left and right index fingers, respectively. Cru-

cially, the task involved a spatial–semantic feature conjunction: to make a correct response, one must maintain the identities of the sounds and their respective locations for comparison with the probe. Participants were instructed to respond as quickly and accurately as possible. The intertrial interval (i.e., the distance between the probe offset and the onset of the scene of the following trial) was jittered from 4000 to 5500 ms (100 ms steps, rectangular distribution).

The retro-cue was a visual stimulus that was either Informative (indicating which object to attend) or not (i.e., Uninformative, maintain all four objects). Informative retro-cues directed attention to either a particular location (Spatial) or to a particular category (Semantic). Spatial retro-cues took the form of a number corresponding to the location of a sound object: 1, -90° ; 2, -30° ; 3, $+30^\circ$; 4, $+90^\circ$; and 5, Uninformative. Semantic retro-cues were presented as a single letter indexing the category of a sound: H, human; A, animal; M, music; O, object; and X, Uninformative. To reduce task switching and confusion, Semantic (letter) and Spatial (number) retro-cues were presented in separate blocks. The Informative retro-cues were always valid; on Spatial cue trials, they were always predictive of the location of the probe, and, on Semantic cue trials, they always indicated the category that would be probed. Importantly, on Informative–Spatial cue trials, participants made a Semantic decision, and on Informative–Semantic cue trials, they made a Spatial decision (Fig. 1). On Uninformative trials, they did their best to remember as many sounds and their respective locations as possible for comparison with the probe sound.

Participants completed a total of 576 trials during EEG recording; there were 192 trials within each of the three retro-cue conditions (i.e., Semantic, Spatial, and Uninformative). For the Uninformative trials, 96 were presented during Spatial blocks (“5”), and 96 were presented during Semantic blocks (“X”). Within each cue condition, half of the trials involved a change (i.e., 96 Match and 96 Non-Match). The location and category of a probe occurred with equal probability within each participant’s session.

Participants completed a pair of practice blocks (one with Spatial retro-cues and one with Semantic retro-cues), followed by eight pairs of experimental blocks (36 trials per block; 72 per pair). The order of spatial/semantic blocks was counterbalanced across participants. Trial order across the experiment was randomized for each participant. Because of one participant’s confession that he did not pay attention to the cues during the first block pair, his first 72 trials were removed from the analysis.

EEG data acquisition and analysis

EEG acquisition. EEG data were acquired using the Neuroscan Synamps2 system (Compumedics) and a 66-channel cap [including four EOG channels, a reference channel during recording (the midline central channel, Cz), and a ground channel (the anterior midline frontal channel, AFz)]. EEG data were digitized at a sampling rate of 500 Hz, with an online bandpass filter from 0.05 to 100 Hz applied.

EEG preprocessing. EEG preprocessing and analyses were performed using a combination of EEGLAB (version 11.4.0.4b; Delorme and Makeig, 2004) and in-house MATLAB (version 7.8) code. Continuous data files were imported into EEGLAB and downsampled to 250 Hz, and the four EOG channels were removed from the data, leaving 60 channels. The continuous data were epoched into segments of -4700 to 1936 ms relative to the retro-cue onset. This long prestimulus period allowed for a 1.5 s period of resting-state brain activity (i.e., the power reference for time–frequency analysis) uncontaminated by either the motor response from the previous trial or the onset of the memory array of the current trial.

To remove artifacts, independent components analysis (ICA) was done, as implemented in EEGLAB. Before the ICA, noisy epochs were initially removed through both automated and visual inspection, resulting in a pruned dataset. The ICA weights resulting from each participant’s pruned dataset were exported and applied to the originally epoched data (i.e., the data containing all epochs, before auto- and visual-rejection of epochs). Subsequently, IC topography maps were inspected to identify and remove artifactual components from the data

(i.e., eyeblinks and saccades). One to three ICs were removed from each participant’s data.

Next, bad channels were interpolated using spherical interpolation, only during the time ranges in which they were bad, and no more than seven channels were interpolated at any given time. On average, 2.4% of the data was interpolated (range across subjects, 0–5.2%). The data were then re-referenced to the average reference, and channel Cz (the reference during recording) was reconstructed. For analysis of event-related potentials (ERPs), the data were baselined to -300 to 0 ms relative to the retro-cue onset; for the time–frequency analysis, the data were baselined to -4700 to -4400 ms. Threshold artifact rejection was used to remove any leftover noisy epochs with deflections exceeding ± 150 μ V from -300 to $+1936$ ms (time-locked to retro-cue onset) for both ERP and time–frequency analyses and also from -4200 to -2700 ms (which served as the power reference period) for time–frequency analysis. For ERP analyses, the data were low-passed filtered at 25 Hz, using a zero-phase finite impulse response least-square filter in MATLAB, and then the epochs were trimmed from -300 to $+1896$ ms. Finally, the trials were sorted according to condition (Semantic, Spatial, Uninformative), and only correct trials were analyzed beyond this point.

Statistical analyses

Behavioral data. Accuracy and response time (RT) were outcome measures for the sound identification and change detection tasks. For the change detection task, one-way ANOVAs with retro-cue condition as the single factor, followed by *post hoc* Tukey’s HSD tests, were performed on each of these outcome measures (i.e., accuracy and RT) using Statistica software (StatSoft). Only RTs from correct trials and those that were 5 s or faster were included in the RT analysis.

ERPs. To isolate orthogonal topographies contributing to the ERP data, we used a spatial principal component analysis (PCA; Spencer et al., 2001). First, we gathered mean voltages for each participant, condition, and channel from 0 to 1896 ms. This data matrix was normalized (z-scored) and entered into the PCA. The first three components (chosen based on a scree plot) were rotated (varimax), and the component scores for each condition, as well as the loadings (coefficients), were obtained. The resulting topographical loadings indicate which channels contribute to its corresponding set of component scores.

This procedure was done three times to examine three contrasts: (1) Informative–Semantic versus Uninformative; (2) Informative–Spatial versus Uninformative; and (3) Informative–Semantic versus Informative–Spatial. (Please note that, for both the ERP–PCA and oscillatory analyses, there were no theoretically or physiologically relevant differences when contrasting Uninformative–Semantic and Uninformative–Spatial trials; thus, we collapsed across these trials and refer to them as “Uninformative.”) Permutation tests, as implemented in EEGLAB [statcond.m, 5000 permutations, $p < 0.05$ false discovery rate (FDR; Benjamini and Yekutieli, 2001)], were used to determine the time points over which the component scores of the contrasted conditions were statistically different. Together, the first two contrasts (Semantic vs Uninformative and Spatial vs Uninformative) show the extent to which domain-general neural correlates of auditory reflective attention are exhibited by both informative conditions, whereas the third contrast (Semantic vs Spatial) isolates domain-specific effects. (Please note that, for the domain-specific contrasts, we switched the labeling of Components 2 and 3, when applicable, so that the domain-specific results could be more easily linked with the domain-general results.)

Oscillatory analysis. Individual time–frequency plots [event-related spectral power (ERSP)] were computed using EEGLAB (newtimef.m). To model ERSP, we used the power from -4200 to -2700 ms as the reference and a three-cycle Morlet wavelet at the lowest frequency (8 Hz); the cycles linearly increased with frequency by a factor of 0.5, resulting in 18.75 cycles at the highest frequency (100 Hz). The window size was 420 ms wide, incremented every ~ 22 ms in the temporal domain and 0.5 Hz in the frequency domain. ERSP indicates the change in signal power across time at each frequency, relative to the reference period, in decibels. We divided the frequency range into six frequency bands: (1) alpha (8–13 Hz); (2) low-beta (13.5–18 Hz); (3) mid-beta (18.5–25 Hz); (4) high-beta (25.5–30 Hz); (5) low-gamma (30.5–70 Hz); and (6) high-

gamma (70.5–100 Hz). Please note that, originally, the time–frequency analysis was conducted with a minimum frequency of 3 Hz, which required a longer window size. However, there were no significant differences among retro-cue conditions in the theta band (4–7 Hz). Because the window size determines how much time at each end of the epoch is “cutoff,” we ran the analysis again, as described above, with 8 Hz as the minimum frequency to examine activity at longer latencies after the retro-cue.

For each time point, spectral power values within each frequency band were averaged within each channel and condition (Uninformative, Informative–Semantic, Informative–Spatial). For each frequency band, permutation tests (5000 permutations; $p < 0.05$ FDR or $p < 0.005$ uncorrected) were done to reveal the time points and channels showing either a domain-general (Informative–Semantic vs Uninformative, Informative–Spatial vs Uninformative) or domain-specific (Informative–Semantic vs Informative–Spatial) effect. The uncorrected oscillatory results were included as cautious findings, which require future studies for verification. To reduce the number of comparisons made, we included only the post-retro-cue time points in the permutation tests, but data across the entire epoch are summarized in the results figures.

Additional ERP and oscillatory analyses: contrasting trials with fast and slow RTs. A secondary goal of the study was to examine how brain activity after an Informative retro-cue might differ according to RT, by blocking Semantic and Spatial cue trials into subsets with either Fast or Slow RTs (all correct responses). Presumably, on Fast RT trials, participants have effectively deployed attention to the retro-cued representation, whereas attention may have remained divided across two or more representations on Slow RT trials. This idea led to two additional analyses. First, we reassessed the domain-specific contrast including only trials with Fast or Slow RTs (i.e., Semantic–Fast vs Spatial–Fast and Semantic–Slow vs Spatial–Slow). Second, we collapsed across Informative cue conditions, leading to a Fast versus Slow contrast.

When selecting Slow and Fast subsets, we matched RTs across Semantic and Spatial conditions. To do this, we sorted each participant’s Semantic and Spatial trials according to the RT of each trial, excluding trials with RTs > 5 s. For each participant, the 30% fastest Semantic trials (+ 1 trial) and 30% slowest Spatial trials (+ 1 trial) were selected as the Semantic–Fast and Spatial–Slow trials. The Semantic–Fast and Spatial–Slow mean RTs were then used to select the Spatial–Fast and Semantic–Slow trials, respectively. For example, the Spatial trial with the RT closest to the Semantic–Fast RT mean, along with the surrounding 30% (15% faster, 15% slower) trials, were selected as the Spatial–Fast trials. Semantic–Slow trials were selected in a similar manner; originally, 15% of trials on each side of the Semantic starting point were chosen, but this led to significantly slower Semantic–Slow RTs than Spatial–Slow RTs. This criterion was adjusted to 18% faster, 12% slower, leading to well matched RTs for the Semantic–Slow and Spatial–Slow conditions. If these percentages exceeded the number of trials available, no adjustment was made and all available trials on that tail of the distribution were collected. This selection procedure was repeated for each participant and was done separately for the ERP and oscillatory analyses.

For the ERPs and oscillatory analyses, there were on average 42.5 trials per condition (range, 29–52) and 37.6 trials per condition (range, 21–52), respectively. For the ERPs, the mean Fast RTs co-occurred with the probe and were 702.6 ms for Semantic and 702.3 ms for Spatial trials; the mean Slow RTs were 1630.2 ms for Semantic and 1660.2 ms for Spatial trials. For the oscillations, the mean Fast RTs were 709.1 ms for Semantic and 712.4 ms for Spatial; the Semantic–Slow mean RT was 1639.8 ms, and the Spatial–Slow was 1653.0 ms. ERP–PCA and oscillatory analyses were rerun with these subsets. We also analyzed ERPs (without PCA), using permutation tests in the same manner as the time–frequency data (5000 permutations; FDR 0.05 or $p < 0.005$ uncorrected).

Neural oscillations–behavior correlations. We examined whether spectral power in the alpha, low-beta, mid-beta, or high-beta frequency bands correlated with task performance. We first calculated a throughput measure of performance (Accuracy/RT; Salthouse and Hedden, 2002) for each participant and condition (Semantic, Spatial, Uninformative). We then averaged the group data across the three retro-cue conditions, separately for each frequency band, and chose the 15 electrodes with the

Figure 2. *A*, Group accuracy results. *B*, Group RT results. Error bars represent within-subjects 95% confidence intervals. R-C, Retro-cue.

largest spectral power desynchronization between 500 and 1000 ms after retro-cue onset to include in the correlations as follows: (1) alpha channels: Oz, O1/2, POz, PO3/4, Pz, P1/2, P3/4, P5/6, P7/8; (2) low-beta channels: O1/2, POz, PO3/4, Pz, P1/2, P3/4, P5/6, P7, CP1, C1; (3) mid-beta channels: POz, PO3/4, Pz, P1/2, P3/4, P5, CPz, CP1/2, Cz, C1/2; and (4) high-beta channels: PO3, Pz, P1/2, P3/4, P5, CPz, CP1/2, CP5, Cz, C1/2, C4. For each participant, condition, and frequency band, we found the mean spectral power, collapsing across the specified time points and channels. For each frequency band and condition, we performed two-tailed hypothesis tests on Pearson’s correlation coefficients.

Results

Behavioral data

At the beginning of the session, participants listened to and reported the category of each sound. This was done to ensure that they could effectively use the Informative–Semantic retro-cues. Participants performed well [Accuracy (i.e., percentage correct), $97.3 \pm 1.2\%$, mean $\pm$ SD; RT, 1358 ± 370 ms, relative to sound onset], indicating that they could readily identify the sounds used in the change detection task.

In the change detection task (Fig. 1), Semantic and Spatial retro-cues were presented in separate blocks of trials; this was done to prevent task switching and confusion. Both block types included Uninformative retro-cue trials. Therefore, we first examined the effect size of the performance differences between the Semantic Uninformative and Spatial Uninformative trials. Semantic (mean Accuracy, 74.6%; mean RT, 1418 ms) and Spatial

Figure 3. *A*, Informative–Semantic versus Uninformative PCA results. *B*, Informative–Spatial versus Uninformative PCA results. Virtual ERPs (i.e., component scores across time for each condition and component) are displayed on the left, and the corresponding PC topographies are displayed on the right. The height of these virtual ERP ribbons denotes the mean component score $\pm$ the within-subjects SE at each time point. In the topographies, warm colors represent channels in which the corresponding component scores are evident (positive loadings), whereas cool colors indicate sites at which the inversion of the component scores are observed (negative loadings). R-C, retro-cue.

(73.4%, 1430 ms) Uninformative retro-cues led to small effect sizes in accuracy (Cohen's $d_z = 0.22$) and RT (Cohen's $d_z = 0.076$), which were sufficiently small to warrant collapsing across Semantic and Spatial Uninformative trials for the subsequent analyses.

Accuracy (i.e., overall percentage correct) and RT data (Fig. 2) were analyzed using one-way ANOVA, followed by Tukey's HSD *post hoc* test. There were significant main effects of retro-cue condition for accuracy ($F_{(2,30)} = 7.12, p = 0.0029, \eta_p^2 = 0.32$) and

RT ($F_{(2,30)} = 36.01, p < 0.00001, \eta_p^2 = 0.71$). For accuracy, Tukey's tests revealed that Spatial retro-cues enhanced performance compared with Uninformative retro-cues ($p = 0.0029$), but there were no significant differences between Semantic and Uninformative cue conditions ($p = 0.23$), nor between Semantic and Spatial retro-cues ($p = 0.11$). For RT, both Semantic ($p < 0.001$) and Spatial ($p < 0.001$) retro-cues led to significantly faster RTs than the Uninformative retro-cues; furthermore, participants responded faster on Spatial than Semantic retro-cue trials ($p = 0.027$). In summary, both Spatial and Semantic retro-cues facilitated RTs relative to Uninformative cue trials, but participants also performed better on Spatial but not Semantic retro-cue trials.

ERPs: spatial PCA

We conducted a spatial PCA on the ERP data to elucidate the domain-general (Informative–Semantic vs Uninformative, Fig. 3*A*; Informative–Spatial vs Uninformative, Fig. 3*B*) and domain-specific (Informative–Semantic vs Informative–Spatial, Fig. 4*A*) neural correlates of attentional orienting to a sound object representation. Using a scree plot, we identified three spatial (topographical) components that explained a total of 62.5% of the variance in the Semantic versus Uninformative contrast, 65.0% of the variance in the Spatial versus Uninformative contrast, and 64.0% of the variance in the Semantic versus Spatial contrast. These three topographical principal components were qualitatively similar for all three contrasts.

Regarding the domain-general PCA results, Components 1 and 2, but not 3, showed significant differences between both Informative cue conditions and Uninformative trials. Component 1, which loaded positively (i.e., was evident) at parieto-occipital electrodes, showed greater positivity for both Semantic and Spatial trials than Uninformative trials between 600 and 620 ms after retro-cue onset and again at ~700–725 ms for Semantic and 700 to 750 ms for Spatial trials. Component 2,

evident in left frontocentral electrodes, also significantly differed between Informative and Uninformative trials. For the Semantic versus Uninformative contrast, there was an early difference at ~145 ms; for the Spatial versus Uninformative contrast, there were early differences at ~110 and 250 ms. These differences likely reflect retro-cue interpretation and will not be discussed further. For both contrasts, there was a left-dominant frontocentral sustained negativity, specific to both types of Informative trials. This sustained potential be-

gan ~ 430 ms and lasted beyond 1800 ms on Semantic trials; on Spatial trials, it occurred from ~ 485 to 1650 ms. Together, these results suggest that two dissociable, domain-general processes, one occurring in parietal/posterior channels and the other observed over the left frontocentral cortex, are involved in orienting attention to a sound object representation.

Now turning to the domain-specific results, only Component 3 showed differences between Informative–Semantic and Informative–Spatial trials. The Semantic trials were more positive than Spatial trials between ~ 570 and 700 ms, whereas the Spatial trials were more positive than Semantic trials starting ~ 1075 ms and extending beyond 1800 ms after retro-cue onset. These modulations were observed over the temporal scalp region (bilaterally), with a right-lateralized reversal in parietal channels. Note that, if the component time course and topography are inverted (multiplied by -1), this earlier difference would be more positive for the Spatial condition and corresponds to activity over the right-lateralized centroparietal cortex. The later, more sustained difference shows a reversal in the polarity of the Semantic and Spatial traces, such that the previously more negative Spatial trace became more positive than the Semantic trace in temporal electrodes.

To further understand these differences, we repeated the domain-specific contrast only including Fast or Slow RTs (matched across Semantic and Spatial trials), as shown in Figure 4, *B* and *C*. Again, three components were extracted that explained 61.2 and 62.3% of the variance in the Semantic versus Spatial contrast for the Fast and Slow subsets, respectively. Only Component 3 showed significant differences for both contrasts, although the pattern of results was strikingly different for the Fast and Slow contrasts. Relative to Spatial–Fast, Semantic–Fast trials showed increased activity in temporal electrodes, from 550 to 875 ms; similar to the primary Semantic versus Spatial PCA, the inversion of this component shows that Spatial–Fast trials led to increased activity over the right centroparietal cortex. However, on Slow RT trials, domain-specific differences were observed in temporal and to some extent frontal channels, with a reversal over the right parietal cortex. Semantic cues led to a greater negativity than Spatial cues at ~ 115 , 230–250, 300–325, and 400 ms after the retro-cue. There was also a sustained difference in temporal/frontal channels from ~ 1050 to 1850 ms that was slightly positive during Spatial–Slow trials and slightly negative during Semantic–Slow trials. Thus, from these results, it appears that the previously described domain-specific effect (exhibited in Fig. 4*A*) occurred predominantly during the Fast trials,

Figure 4. Domain-specific results of the Semantic versus Spatial PCA (Component 3) using all correct, artifact-free trials (*A*), correct trials with Fast RTs (matched across conditions) (*B*), and correct trials with Slow RTs (matched across conditions) (*C*). The height of the ribbons in *A–C* reflect the mean component score $\pm$ the within-subjects SE at each time point. The corresponding PC topographies are displayed to the right of the virtual ERPs. *D* shows Semantic–Fast versus Spatial–Fast and Semantic–Slow versus Spatial–Slow ERP results from a few channels with strong loadings in the PCA, namely FT9, FT10, and CP2 (marked by darkened electrodes in the topographies of *B* and *C*). In channels FT9 and CP2, notice the earlier domain-specific modulation (~ 700 ms) that is specific to trials with Fast RTs, followed by the later differences observed in CP2 for the Slow contrast and FT10 for the Fast contrast. Sem, Semantic; Spa, spatial; unc, uncorrected.

whereas the later difference was driven primarily by the Slow trials. However, from examination of the ERPs (Fig. 4*D*), it was clear that, on Fast trials, Spatial cues led to stronger activity in right temporal channels than Semantic cues, from ~ 1200 to 1550 ms. This later effect may reflect memory maintenance of the retro-cued sound identity on Spatial retro-cue trials, in anticipation of the semantic decision. These results suggest that attention can be deployed in a feature-specific manner to an auditory representation, specifically on the Fast RT trials, in which participants were able to successfully orient attention to the retro-cued representation.

Neural oscillation results

The oscillatory results are displayed in Figures 5 (domain-general results), 6 (domain-specific results), 7 (Fast vs Slow RT Informative cue trials), and 8 (oscillatory activity–behavior correlations). Because gamma activity did not show reliable differences between conditions, the results focus on alpha (8–13 Hz), low-beta

Figure 5. Domain-general oscillations results for the alpha (**A**), low-beta (**B**), and mid-beta (**C**) bands. For each frequency band, the power time course in one channel is displayed; the height of the ribbons indicates the mean power $\pm$ the within-subjects SE at each time point. Within each power time course plot, the significance markings toward the top indicate time points that were significantly different between Semantic and Uninformative (Red), Spatial and Uninformative (Blue), or both (Yellow). Significance values for both $p < 0.05$ FDR-corrected and $p < 0.005$ uncorrected are displayed. The scalp topography maps at the bottom left show mean power in each condition and band from 416 to 916 ms. unc, Uncorrected; Uninf, uninformative.

(13.5–18 Hz), mid-beta (18.5–25 Hz), and high-beta (25.5–30 Hz) oscillations.

The domain-general contrasts (Semantic vs Uninformative and Spatial vs Uninformative; Fig. 5) revealed widespread power suppression [i.e., event-related desynchronization (ERD)] in the alpha and low-beta bands, peaking at ~ 700 ms and largest in posterior electrodes. The domain-general effects tended to be stronger and longer-lasting for the Semantic (alpha, ~ 400 –1000 ms in frontal and posterior channels and extending to 1400 ms especially in left parietal channels, $p < 0.005$, uncorrected; low-beta, ~ 400 –1000 ms in posterior channels and ~ 1100 –1400 ms especially in left posterior and right central channels, $p < 0.05$ FDR) than the Spatial contrast (alpha, ~ 500 –1000 ms; low-beta, ~ 500 –900 ms; both strongest in posterior channels, $p < 0.005$ uncorrected). [Note that the FDR threshold (~ 0.01) was actually more lenient than the uncorrected threshold of 0.005 for the alpha Semantic contrast, so the described effects were based on the uncorrected threshold.] Furthermore, both Semantic and Spatial retro-cues resulted in significantly stronger ERD than Uninformative retro-cues from ~ 600 to 800 ms in central and parietal channels for mid-beta and from 700 to 750 ms in right centroparietal electrodes for high-beta (both $p < 0.005$, uncorrected). Also, relative to Uninformative cues, Semantic cues led to stronger ERD over the centroparietal cortex between ~ 1100 and 1400 ms for the mid-beta band ($p < 0.05$ FDR or $p < 0.005$ uncorrected) and between ~ 1150 and 1250 ms for the high-beta band ($p < 0.005$ uncorrected). In summary, both types of Informative cues led to stronger ERD surrounding ~ 700 ms compared with Uninformative cues, suggesting a domain-general correlate of orienting attention within auditory STM, but only

Semantic retro-cues resulted in stronger ERD later, from ~ 1100 to 1400 ms.

Domain-specific differences (Informative-Semantic vs Informative-Spatial retro-cues; Fig. 6) generally revealed stronger ERD after Semantic compared with Spatial retro-cues. In the alpha and low-beta bands, significant domain-specific modulations were evident early (centered at ~ 148 ms, $p < 0.05$ FDR) in frontal and lateral parietal/occipital electrodes and later (~ 1100 –1300 ms, $p < 0.05$ FDR or $p < 0.005$ uncorrected) in frontal and (especially left) lateral parietal/posterior channels. Stronger mid- and high-beta (~ 1200 –1700 ms) ERD occurred after Semantic, compared with Spatial, retro-cues in centroparietal electrodes ($p < 0.005$ uncorrected). During examination of the domain-specific RT contrasts (results not shown), the Fast, but not Slow, contrast showed similar domain-specific alpha, low-beta, and mid-beta results ($p < 0.005$ uncorrected) as those in the original domain-specific analysis. This suggests that efficacious feature-based attentional orienting underlay these domain-specific modulations. However, Slow RT trials seemed to drive the domain-specific high-beta results more so than the Fast RT trials; this implies that a more general feature-based process may underlie high-beta activity. Except for the early domain-specific activity, the domain-general and later domain-specific ERD results appeared to reflect induced activity (i.e., not phase-locked to the retro-cue onset) except for a phase-locked response from 650 to 750 ms in the alpha and low-beta bands; this interpretation is based on inspection of the corresponding intertrial phase coherence spectrograms (data not shown).

Figure 6. Domain-specific oscillatory results are displayed for the alpha (A), low-beta (B), mid-beta (C), and high-beta (D) bands. The power time course is illustrated for one channel per band, with markings indicating time points that significantly differed between Semantic and Spatial conditions at $p < 0.005$ uncorrected and, when applicable, at $p < 0.05$ FDR. The scalp topographies (bottom left) depict the mean power from 1288 to 1624 ms for each condition and band, as well as the power at 148 ms for the alpha band. unc, Uncorrected.

Fast RT versus Slow RT results

Next, we examined differences between Fast and Slow RT trials, collapsed across the Informative–Semantic and Informative–Spatial retro-cues (Fig. 7). The previously observed ERD peak in the alpha, low-beta, and mid-beta bands modulated RT, such that stronger ERD preceded Fast, relative to Slow, RTs. In the alpha band, this effect was evident between ~300 and 900 ms in frontal and parietal/posterior channels ($p < 0.05$ FDR or $p < 0.005$ uncorrected). Similarly, in the low-beta band, there were significant differences between ~300 and 800 ms in central, parietal, and posterior channels ($p < 0.005$ uncorrected). For the mid-beta band, there were small differences in central channels (600–700 ms, $p < 0.005$ uncorrected). Using a spatial PCA, the ERPs revealed a small difference at ~1830 and 1890 ms in a component resembling the Semantic versus Spatial Component 3 topography (Fig. 4A), in which Fast trials had increased activity relative to Slow trials over temporal sites ($p < 0.05$ FDR); this effect may reflect greater anticipation of the probe on Fast RT trials. However, during examination of the individual ERPs, there was a longer-lasting modulation in which a stronger left frontal sustained negativity, starting at ~600 ms, was present on Fast compared with Slow trials ($p < 0.005$ uncorrected; Fig. 7D).

Neural oscillation–behavior correlations

Finally, we examined whether alpha and beta ERD were related to task performance, as a recent model would predict (Hanslmayr et al., 2012). We correlated a throughput measure of performance (Accuracy/RT) with alpha, low-, mid-, and high-beta power in each condition (Informative–Semantic, Informative–Spatial,

and Uninformative). Figure 8 shows the results for all four frequency bands. Both alpha and low-beta power significantly negatively correlated with performance (i.e., better performance, stronger ERD) for the Informative–Semantic (alpha, $r = -0.50$, $p = 0.048$; low-beta, $r = -0.60$, $p = 0.014$) and Informative–Spatial (alpha, $r = -0.56$, $p = 0.025$; low-beta, $r = -0.60$, $p = 0.014$) conditions. In the Uninformative condition, the correlation was not significant for the alpha band ($r = -0.41$, $p = 0.12$) but was nearly significant for low-beta ($r = -0.50$, $p = 0.051$). There were also significant correlations for the mid- and high-beta bands for the Informative–Semantic (mid-beta: $r = -0.63$, $p = 0.0083$; high-beta: $r = -0.68$, $p = 0.0038$) and Uninformative (mid-beta, $r = -0.51$, $p = 0.045$; high-beta, $r = -0.50$, $p = 0.049$) conditions but not for the Informative–Spatial condition (mid-beta, $r = -0.45$, $p = 0.080$; high-beta, $r = -0.34$, $p = 0.20$).

Discussion

This study reveals neural correlates of attentional orienting to auditory representations in STM and, in doing so, provides new evidence pertaining to the object-based account of auditory attention. Both types of Informative retro-cues facilitated performance (although Spatial cues to a greater extent than Semantic) relative to Uninformative retro-cues, demonstrating that an auditory representation can be attended through its semantic or spatial feature, extending the results of a previous behavioral study (Backer and Alain, 2012).

The PCA of ERP data revealed two components that reflected domain-general attentional orienting, putatively involving the left frontocentral cortex (~450–1600+ ms) and bilateral poste-

Figure 7. This figure illustrates that Informative retro-cue trials with Fast RTs (collapsed across cue type) modulated alpha (**A**), low-beta (**B**), mid-beta (**C**), and ERP (**D**) activity relative to trials with Slow RTs. In each panel, the time course from one channel is displayed; this channel is darkened in the corresponding topographies. The time points that significantly differed between Fast and Slow are marked toward the top of the plots for the oscillations and toward the bottom of the plot for the ERPs. uncorr., Uncorrected.

rior cortex (~600–750 ms), mirroring the commonly observed frontoparietal attentional network. The left frontal activity likely reflects executive processing involved in the deployment of and sustained attention to the retro-cued representation. Bosch et al. (2001) retro-cued participants to visually presented verbal and nonverbal stimuli and found a similar sustained left frontal negativity that they interpreted as a domain-general correlate of attentional control, an idea supported by fMRI studies showing prefrontal cortex activation during reflective attention (Nobre et al., 2004; Johnson et al., 2005; Nee and Jonides, 2008). After the frontal activation, the posterior positivity may be mediating attentional biasing to the retro-cued representation (Grent-‘t-Jong and Woldorff, 2007).

We also found domain-general oscillatory activity. Both Semantic and Spatial retro-cues led to stronger ERD (i.e., power suppression) than Uninformative retro-cues from ~400 to 1000 ms after retro-cue onset. This alpha/beta ERD coincided with the sustained frontocentral ERP negativity; previous studies have suggested that event-locked alpha activity can underlie sustained ERPs (Mazaheri and Jensen, 2008; van Dijk et al., 2010). Alpha/beta ERD is thought to reflect the engagement of local, task-relevant processing (Hanslmayr et al., 2012) and has been implicated in studies involving auditory attention to external stimuli (McFarland and Cacace, 2004; Mazaheri and Picton, 2005; Kerlin et al., 2010; Banerjee et al., 2011), auditory STM retrieval (Krause et al., 1996; Pesonen et al., 2006; Leinonen et al., 2007), and visual memory retrieval (Zion-Golumbic et al., 2010; Khader and Rösler, 2011). Because memory retrieval is a form of reflective attention (Chun and Johnson, 2011) and because attentional control to external stimuli and internal representations has

been shown to involve overlapping neural processes (Griffin and Nobre, 2003; Nobre et al., 2004), it is perhaps not surprising that many studies involving attentional control, including the present one, reveal alpha and/or beta ERD. Furthermore, Mazaheri et al. (2014), using a cross-modal attention task, showed that alpha/low-beta ERD localized to the right supramarginal gyrus during attention to auditory stimuli; this region may also be the source of the ERD in the present study.

Recently, Hanslmayr et al. (2012) proposed the “information by desynchronization hypothesis,” which postulates that a certain degree of alpha/beta desynchronization is necessary for successful long-term (episodic) memory encoding and retrieval. Their hypothesis predicts that the richness/amount of information that is retrieved from long-term memory is proportional to the strength of alpha/beta desynchronization. In the present study, participants with stronger alpha and low-beta ERD tended to perform better than those with weaker ERD on Semantic and Spatial retro-cue trials, suggesting that alpha/low-beta ERD may represent a supramodal correlate of effective attentional orienting to memory. There were also significant correlations between performance and mid- and high-beta ERD 500–1000 ms after Semantic and Uninformative, but not Spatial, retro-cues. This suggests that mid/high-beta may be specifically involved in semantic-based processing, because both Semantic and Uninformative retro-cues involved attending to the semantic feature of the sound representations, especially during this portion of the epoch.

Furthermore, across Informative retro-cue conditions, differences in domain-general activity preceded RT differences. On Fast RT trials, there was stronger alpha/beta ERD, as well as a

Figure 8. Correlations between performance and alpha (**A**), low-beta (**B**), mid-beta (**C**), and high-beta (**D**) power after Informative–Semantic, Informative–Spatial, and Uninformative retro-cues. Each circle represents a participant ($n = 16$). R-C, Retro-cue.

larger left frontocentral sustained negativity than Slow RT trials. Presumably, on Fast RT trials, participants have quickly deployed attention to the retro-cued representation. Consequently, when the probe sound plays, they can respond once the attended representation is compared with the probe. However, on Slow RT trials, participants may have had difficulty orienting attention to the retro-cued representation (e.g., because of fatigue or decreased salience compared with the other sounds), resulting in divided attention across several representations (similar to the Uninformative cue) and more time to complete the comparison process with the probe. These domain-general modulations may reflect divided versus selective attention within auditory STM and corroborate the notion of the object-based account that attention can be selectively oriented to an auditory representation. However, differences in overall attentional state between Fast and Slow trials may have contributed to these neural modulations.

Both oscillatory and PCA results revealed that attention to an auditory representation involves feature-specific processing. When comparing domain-specific (Informative–Semantic vs Informative–Spatial) oscillatory activity, stronger alpha/low-beta ERD was evident early after Semantic retro-cues and appeared to begin after the presentation of the memory array and before the

Semantic retro-cue. Previous studies have shown that semantic processing relies on alpha and beta ERD (Klimesch et al., 1997; Krause et al., 1999; Röhm et al., 2001; Hanslmayr et al., 2009; Shahin et al., 2009). Because Semantic and Spatial retro-cues were blocked separately, participants may have used different strategies during the encoding and rehearsal of the sounds before the retro-cue, resulting in stronger alpha/beta ERD on Informative–Semantic retro-cue trials. Later, stronger alpha and beta ERD was evident after 1100 ms in centroparietal electrodes on Informative–Semantic, relative to Informative–Spatial, retro-cue trials, suggesting that semantic-based reflective attention may take longer than spatial-based orienting. This is consistent with our finding that Informative–Semantic cue trials resulted in longer RTs than Informative–Spatial cue trials and with a previous study showing stronger, longer-latency alpha/beta ERD during auditory semantic, relative to pitch (a low-level feature), processing (Shahin et al., 2009). However, after reexamining the domain-specific oscillations using RT-matched trials, at least the alpha and low-beta effects were specific to the Fast trials, when attention was successfully oriented to the retro-cued representation. This suggests that domain-specific differences during attentional orienting likely drove these effects. However, because the

RT followed the probe, we cannot definitively determine whether RT differences between Semantic and Spatial conditions arose during attentional orienting, the probe decision, or both; consequently, despite using matched RTs, RT/difficulty differences may have contributed to these domain-specific differences.

The domain-specific PCA effect was evident in temporal channels, with an inversion in right-lateralized centroparietal sites. Although we cannot be sure of the precise anatomical locations underlying this activity, we can conjecture that these effects are in line with the dual-pathway (what/where) model (Rauschecker and Tian, 2000; Alain et al., 2001). Following this model, we can surmise that the earlier feature-specific difference, showing a positive peak in temporal electrodes, reflects access to an auditory representation using its semantic feature. Simultaneously, activity after Spatial cues was more evident over right-lateralized centroparietal channels, again presumably reflecting dominance of spatial processing. Importantly, these domain-specific effects, peaking between 500 and 700 ms, were evident for Semantic and Spatial trials with Fast, but not Slow, RTs, supporting the idea that successful attentional orienting to the retro-cued representation drove the observed domain-specific modulations.

A later domain-specific difference between ~1100 and 1700 ms also emerged, in which activity after a Spatial retro-cue was more positive in temporal sites than that after a Semantic retro-cue. However, this was primarily driven by a sustained frontoparietal difference in the Slow RT contrast, which may reflect greater sustained cognitive effort during the Slow–Semantic trials. Although this crossover effect was not apparent in the Fast contrast PCA, it was evident in the ERPs, possibly reflecting attention to the semantic feature (i.e., sound identity) in the latter half of the silent delay after a Spatial retro-cue. This highlights the possibility that attention to one feature within a sound representation can be temporally dissociated from attention to the other feature(s) of that representation.

In summary, we have revealed domain-general and domain-specific neural dynamics involved in orienting attention to one of several coexisting sound representations in STM. These results support the assumptions of the object-based account that concurrent sound objects can coexist in STM and that attention can be deployed to one of several competing objects. However, we found some feature-specific modulations involved in the deployment of attention within auditory STM, providing evidence that attention to one feature may not result in attentional capture of the other features of the object (Krumbholz et al., 2007), as a strict formulation of the object-based account might assume. However, the degree to which domain specificity is observed may depend on the types of features that are attended. For example, retro-cuing attention to two nonspatial acoustic features may result in a lack of domain-specific neural modulations. Future studies should investigate this idea, furthering our knowledge of how various stimulus features are represented in auditory STM.

References

- Alain C, Arnott SR (2000) Selectively attending to auditory objects. *Front Biosci* 5:D202–D212. [CrossRef Medline](#)
- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) “What” and “where” in the human auditory system. *Proc Natl Acad Sci U S A* 98:12301–12306. [CrossRef Medline](#)
- Alain C, Schuler BM, McDonald KL (2002) Neural activity associated with distinguishing concurrent auditory objects. *J Acoust Soc Am* 111:990–995. [CrossRef Medline](#)
- Backer KC, Alain C (2012) Orienting attention to sound object representations attenuates change deafness. *J Exp Psychol Hum Percept Perform* 38:1554–1566. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29:1165–1188. [CrossRef](#)
- Bosch V, Mecklinger A, Friederici AD (2001) Slow cortical potentials during retention of object, spatial, and verbal information. *Brain Res Cogn Brain Res* 10:219–237. [CrossRef Medline](#)
- Chun MM, Johnson MK (2011) Memory: enduring traces of perceptual and reflective attention. *Neuron* 72:520–535. [CrossRef Medline](#)
- Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134:9–21. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Dyson BJ, Ishfaq F (2008) Auditory memory can be object based. *Psychon Bull Rev* 15:409–412. [CrossRef Medline](#)
- Grent-’t-Jong T, Woldorff MG (2007) Timing and sequence of brain activity in top-down control of visual-spatial attention. *PLoS Biol* 5:e12. [CrossRef Medline](#)
- Griffin IC, Nobre AC (2003) Orienting attention to locations in internal representations. *J Cogn Neurosci* 15:1176–1194. [CrossRef Medline](#)
- Hanslmayr S, Spitzer B, Bäuml KH (2009) Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cereb Cortex* 19:1631–1640. [CrossRef Medline](#)
- Hanslmayr S, Staudigl T, Fellner MC (2012) Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Front Hum Neurosci* 6:74. [CrossRef Medline](#)
- Hill KT, Miller LM (2010) Auditory attentional control and selection during cocktail party listening. *Cereb Cortex* 20:583–590. [CrossRef Medline](#)
- Johnson MK, Raye CL, Mitchell KJ, Greene EJ, Cunningham WA, Sanislow CA (2005) Using fMRI to investigate a component process of reflection: prefrontal correlates of refreshing a just-activated representation. *Cogn Affect Behav Neurosci* 5:339–361. [CrossRef Medline](#)
- Kerlin JR, Shahin AJ, Miller LM (2010) Attentional gain control of ongoing cortical speech representations in a “cocktail party.” *J Neurosci* 30:620–628. [CrossRef Medline](#)
- Khader PH, Rösler F (2011) EEG power changes reflect distinct mechanisms during long-term memory retrieval. *Psychophysiology* 48:362–369. [CrossRef Medline](#)
- Klimesch W, Doppelmayr M, Pachinger T, Russegger H (1997) Event-related desynchronization in the alpha band and the processing of semantic information. *Brain Res Cogn Brain Res* 6:83–94. [CrossRef Medline](#)
- Krause CM, Lang AH, Laine M, Kuusisto M, Pörn B (1996) Event-related EEG desynchronization and synchronization during an auditory memory task. *Electroencephalogr Clin Neurophysiol* 98:319–326. [CrossRef Medline](#)
- Krause CM, Åström T, Karrasch M, Laine M, Sillanmäki L (1999) Cortical activation related to auditory semantic matching of concrete versus abstract words. *Clin Neurophysiol* 110:1371–1377. [CrossRef Medline](#)
- Krumbholz K, Eickhoff SB, Fink GR (2007) Feature- and object-based attentional modulation in the human auditory “where” pathway. *J Cogn Neurosci* 19:1721–1733. [CrossRef Medline](#)
- Leinonen A, Laine M, Laine M, Krause CM (2007) Electrophysiological correlates of memory processing in early Finnish-Swedish bilinguals. *Neurosci Lett* 416:22–27. [CrossRef Medline](#)
- Mazaheri A, Jensen O (2008) Asymmetric amplitude modulations of brain oscillations generate slow evoked responses. *J Neurosci* 28:7781–7787. [CrossRef Medline](#)
- Mazaheri A, Picton TW (2005) EEG spectral dynamics during discrimination of auditory and visual targets. *Brain Res Cogn Brain Res* 24:81–96. [CrossRef Medline](#)
- Mazaheri A, van Schouwenburg MR, Dimitrijevic A, Denys D, Cools R, Jensen O (2014) Region-specific modulations in oscillatory alpha activity serve to facilitate processing in the visual and auditory modalities. *Neuroimage* 87:356–362. [CrossRef Medline](#)
- McFarland DJ, Cacace AT (2004) Separating stimulus-locked and unlocked components of the auditory event-related potential. *Hear Res* 193:111–120. [CrossRef Medline](#)

- Nee DE, Jonides J (2008) Neural correlates of access to short-term memory. *Proc Natl Acad Sci U S A* 105:14228–14233. [CrossRef Medline](#)
- Nobre AC, Coull JT, Maquet P, Frith CD, Vandenberghe R, Mesulam MM (2004) Orienting attention to locations in perceptual versus mental representations. *J Cogn Neurosci* 16:363–373. [CrossRef Medline](#)
- Pesonen M, Björnberg CH, Hämäläinen H, Krause CM (2006) Brain oscillatory 1–30 Hz EEG ERD/ERS responses during the different stages of an auditory memory search task. *Neurosci Lett* 399:45–50. [CrossRef Medline](#)
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800–11806. [CrossRef Medline](#)
- Röhm D, Klimesch W, Haider H, Doppelmayr M (2001) The role of theta and alpha oscillations for language comprehension in the human electroencephalogram. *Neurosci Lett* 310:137–140. [CrossRef Medline](#)
- Salthouse TA, Hedden T (2002) Interpreting reaction time measures in between-group comparisons. *J Clin Exp Neuropsychol* 24:858–872. [CrossRef Medline](#)
- Shahin AJ, Picton TW, Miller LM (2009) Brain oscillations during semantic evaluation of speech. *Brain Cogn* 70:259–266. [CrossRef Medline](#)
- Shinn-Cunningham BG (2008) Object-based auditory and visual attention. *Trends Cogn Sci* 12:182–186. [CrossRef Medline](#)
- Spencer KM, Dien J, Donchin E (2001) Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology* 38:343–358. [CrossRef Medline](#)
- van Dijk H, van der Werf J, Mazaheri A, Medendorp WP, Jensen O (2010) Modulations in oscillatory activity with amplitude asymmetry can produce cognitively relevant event-related responses. *Proc Natl Acad Sci U S A* 107:900–905. [CrossRef Medline](#)
- Zatorre RJ, Mondor TA, Evans AC (1999) Auditory attention to space and frequency activates similar cerebral systems. *Neuroimage* 10:544–554. [CrossRef Medline](#)
- Zion-Golumbic E, Kutas M, Bentin S (2010) Neural dynamics associated with semantic and episodic memory for faces: evidence from multiple frequency bands. *J Cogn Neurosci* 22:263–277. [CrossRef Medline](#)