

# Attentional Gain Control of Ongoing Cortical Speech Representations in a “Cocktail Party”

Jess R. Kerlin,<sup>1</sup> Antoine J. Shahin,<sup>1,3</sup> and Lee M. Miller<sup>1,2</sup>

<sup>1</sup>Center for Mind and Brain and <sup>2</sup>Department of Neurobiology, Physiology, and Behavior, University of California, Davis, California 95618, and <sup>3</sup>Ohio State University Eye and Ear Institute and Department of Otolaryngology-Head and Neck Surgery, Columbus, Ohio 43212

Normal listeners possess the remarkable perceptual ability to select a single speech stream among many competing talkers. However, few studies of selective attention have addressed the unique nature of speech as a temporally extended and complex auditory object. We hypothesized that sustained selective attention to speech in a multitalker environment would act as gain control on the early auditory cortical representations of speech. Using high-density electroencephalography and a template-matching analysis method, we found selective gain to the continuous speech content of an attended talker, greatest at a frequency of 4–8 Hz, in auditory cortex. In addition, the difference in alpha power (8–12 Hz) at parietal sites across hemispheres indicated the direction of auditory attention to speech, as has been previously found in visual tasks. The strength of this hemispheric alpha lateralization, in turn, predicted an individual’s attentional gain of the cortical speech signal. These results support a model of spatial speech stream segregation, mediated by a supramodal attention mechanism, enabling selection of the attended representation in auditory cortex.

## Introduction

Listening to one person speaking among many others serves as a model for selective attention in ecological environments, as most people depend upon it daily for social interaction and well-being (Pichora-Fuller and Singh, 2006; Shinn-Cunningham and Best, 2008). Commonly referred to as the “cocktail party effect” (Cherry, 1953), this perceptual feat has been studied extensively for decades using behavioral methods (Broadbent, 1957; Treisman and Geffen, 1967; Driver, 2001). Neural evidence for auditory attentional modulation first arose in electroencephalography (EEG), in which the primary approach has been to characterize differences in the transient event-related potentials (ERPs) of attended versus unattended sounds (Picton and Hillyard, 1974; Näätänen and Michie, 1979; Woods et al., 1993; Alcaini et al., 1995). While this approach has expanded our knowledge of auditory attention, with few exceptions (Teder et al., 1993; Coch et al., 2005; Nager et al., 2008) it tends to present sounds in isolation rather than concurrently and therefore may not engage selective attention in the same manner as under the intense perceptual load of a “cocktail party” (Lavie, 1995). It is also limited by treating all sounds, phonemes, or words as discrete events with a stereotyped neural onset response.

Recent EEG and magnetoencephalography (MEG) studies have used novel methods to measure the continuous responses,

rather than stereotyped onset components, to natural speech from early auditory cortex. These signals appear to be closely related to the slow (2–20 Hz) acoustic envelope of speech (Ahissar et al., 2001; Purcell et al., 2004; Abrams et al., 2008; Aiken and Picton, 2008; Lalor et al., 2009) and can differentiate responses to vowels, words, and sentences (Suppes et al., 1997, 1998, 1999; Luo and Poeppel, 2007; Bonte et al., 2009). Given its success in classifying natural speech when presented alone, a continuous neural measure should be especially suited to characterize how selective attention acts on concurrent streams of continuous speech.

Finally, it remains unclear which top-down neural signals are responsible for mediating attention to continuous speech in space. Recent evidence suggests visuospatial attention, in particular the suppression of distracting speakers’ faces, affects comprehension of an attended speaker (Senkowski et al., 2008). Visuospatial attention is known to be associated with relative contralateral alpha suppression at posterior sites, which has been attributed to parietal and/or occipital cortices (Worden et al., 2000; Gruber et al., 2005; Medendorp et al., 2007; Palva and Palva, 2007). This activity, in turn, predicts successful visual detection (Thut et al., 2006). Although it has been proposed that visuospatial attention and auditory spatial attention share an overlapping mechanism, it is unknown whether contralateral posterior alpha suppression occurs or plays a role in auditory perception.

In the current study, we presented two different sentences to listeners, either one sentence at a time or simultaneously with one on each side. While listeners performed a comprehension task, we recorded high-density EEG, filtered into frequency bands ranging from very low (1–4 Hz) to ultra-high (120–160 Hz) frequencies. We hypothesized that selective attention would act via a gain increase on the lower frequencies in auditory cortical

Received July 27, 2009; revised Oct. 23, 2009; accepted Nov. 23, 2009.

This research was supported by a grant from the National Institutes of Health—National Institute on Deafness and Other Communication Disorders (R01-DC008171). We thank Kevin Hill, Kristina Backer, and Chris Bishop for advice and support in data collection, as well as Terry Picton, Ali Mazaheri, Tom Campbell, and Risa Sawaki for advice and technical expertise.

Correspondence should be addressed to Jess R. Kerlin, Center for Mind and Brain, University of California, Davis, 267 Cousteau Place, Davis, CA 95618. E-mail: jrkerlin@ucdavis.edu.

DOI:10.1523/JNEUROSCI.3631-09.2010

Copyright © 2010 the authors 0270-6474/10/300620-09\$15.00/0

activity for the attended sentence, as measured in single trials. We also tested whether lateralization of alpha activity at posterior sites could predict the gain of the attended signal across individuals, implying a mechanistic link between spatial and selective attention in a “cocktail party.”

## Materials and Methods

**Participants.** Fourteen volunteers (8 female) between 18 and 36 years old participated in the experiment. All participants were right-handed native English speakers with normal hearing, no history of neurological problems, and no use of psychoactive medications or drugs in the past month. Participants gave written informed consent in accordance with procedures approved by the University of California Institutional Review Board and were paid for their participation. A single participant was removed from all analyses in the current study, based on near-chance overall behavioral performance (55% accuracy). This was more than three SDs below the accuracy of the group, qualifying as an outlier.

**Speech stimuli.** All speech stimuli were recorded in a sound-dampened chamber from a 25-year-old male speaker. The speech stimuli consisted of two incomplete sentences (sentence A: “Brandon’s wife seemed . . .”; sentence B: “His other friend got the . . .”) of the same duration (1.36 s) and 128 ending words (64 adjectives and 64 nouns). Ending words were selected from the MRC Psycholinguistic Database ([http://www.psy.uwa.edu.au/mrcdatabase/uwa\\_mrc.htm](http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm)). All adjectives were screened to have a familiarity score of  $>500$  and all nouns a concreteness score of  $>500$ . Both nouns and adjectives were required to have 2 syllables. The words were then further selected by a native English speaker to ensure the adjectives and nouns were grammatically correct and semantically plausible as a final word for sentences A and B, respectively.

To perform an analysis of the fundamental frequency of speech, the speech stimuli were further processed in Praat (<http://www.praat.org>) to flatten and alter the fundamental frequency of all speech sounds uniformly. From an average fundamental frequency of 128 Hz for the original speaker, two sets of flattened speech stimuli with fundamentals at 123 Hz and 133 Hz were produced. The resulting sentences and words remained highly intelligible, yet sounded monotonic and were lacking in prosody. This conditioning of the stimulus was intended to produce a frequency-following response (FFR) at the fundamental frequency in the EEG waveform under all conditions. However, we had far fewer trials than previous studies of the FFR (Krishnan, 2002; Dajani et al., 2005; Musacchia et al., 2007), and the vowels of our ongoing speech stimuli were short and discontinuous, which may have led to our lack of observed FFR. This manipulation will therefore not be addressed further in this paper.

Head-related transfer functions (HRTFs), recorded from AuSIM in-the-canal microphones at  $-45^\circ$  (left of midline),  $+45^\circ$  (right of midline), and  $0^\circ$  (midline) along the horizontal azimuth, were obtained for each subject. Each individual’s HRTF was used to filter the speech stimuli so talkers were perceived in virtual external space (Langendijk and Bronkhorst, 2000). Finally, each speech stimulus was normalized to have equal root mean square (rms) amplitudes at a volume of  $\sim 70$  dB hearing level. All speech stimuli were presented with Etymotic ER-4B insert earphones, shielded with grounded metallic tape to avoid transduction artifacts in the EEG recordings. Fifty percent of the sentence waveforms were randomly inverted, with no perceptual consequence, to further preclude any possibility of artifactual phase locking in the EEG signal.

**Presentation and trial structures.** Each trial presented one of three conditions (Single Talker, Selective Attention, and Central Control), each with three possible cue instructions. The Single Talker condition included presenting a sentence immediately followed by an ending word. The ending word was equally likely to be grammatically congruent (e.g., “Brandon’s wife seemed friendly.”) or incongruent (e.g., “Brandon’s wife seemed lizard.”). Single Talker sentences were always presented to either the left or right of the participant. The Selective Attention condition included the simultaneous presentation of different Single Talker sentences to the left and right of the participant

(always on opposite sides). The Central Control condition included the simultaneous presentation of two sentences in the same midline location.

Participants were instructed to attend to the subsequently presented speech stimulus based on one of three cues: a left arrow (“ $<$ ”), a right arrow (“ $>$ ”), or the numeral zero (“0”). For the Single Talker and Selective Attention conditions, the left or right arrow indicated the participant was to attend to the speech presented in that direction. In the Central Control condition, the participants were instructed to attend to the talker based on pitch. During the Central Control condition, participants were therefore told to attend to the lower pitch when given the left cue and to the higher pitch when given the right cue. For seven participants, all speech stimuli in the Single Talker and Selective Attention conditions presented to the left had the fundamental frequency ( $f_0$ ) flattened to 123 Hz, and all speech stimuli presented to the right had  $f_0$  flattened to 133 Hz during the entire session. For the other seven participants, the locations of flattening and the corresponding pitch cue instructions were reversed. In the Central Control condition, participants were required to segregate the speech stimuli using the very small pitch differences present in the Selective Attention condition, which, in contrast, also included strong spatial information. Thus, it served as a behavioral control for whether participants could be using the pitch information in the Selective Attention task, rather than the spatial information.

In all conditions cued with an arrow, participants were told to press the 1 key on the response pad if the attended sentence was grammatically congruent and the 2 key if it was incongruent, regardless of its semantic likelihood. Participants were asked to respond as quickly and accurately as possible. For all subjects and sentence presentation types, participants were told that the “0” cue indicated a passive trial, when they should ignore all speech signals and give no response. For all conditions there were equal numbers of each cue. For Single Talker presentations, the arrow cues always pointed in the direction in which the speech stimulus would actually be presented.

Each trial started with a cue, replaced 1000 ms later by a crosshair that was maintained throughout the rest of the trial, followed 1000 ms later by the onset of the sentence. After the 1364 ms presentation of the sentence, a 1900 ms window was given in which the participant’s response could be counted as accurate. The next trial would start jittered uniformly between 1000 and 2000 ms after the end of the response window (Fig. 1).

Each participant completed 12 blocks of 80 trials, with a short break between each block, for a total of 960 trials, except for two participants who completed 9 blocks, for a total of 720 trials. Forty percent of trials presented Single Talker sentences, 40% presented Selective Attention sentences, and 20% presented Central Control sentences. Trials were presented in random order from a full factorial design that counterbalanced side of presentation, sentence type, and grammar congruency.

**Data acquisition and analysis.** EEG was recorded from a 128-electrode cap continuously throughout trial presentation with the BioSemi ActiveTwo data acquisition system. All recordings were conducted in a sound-dampened, electrically shielded room. The data were originally recorded at a sampling rate of 2048 Hz and subsequently downsampled to a rate of 512 Hz. For some participants, some electrode channels would be visibly, excessively noisy for the entire duration of recording and were therefore marked and interpolated with surrounding electrode sites (Mean number of bad channels = 2.1; SD = 2.7).

All data analysis, except source localization, was performed in MATLAB using a combination of the FieldTrip MATLAB toolbox (<http://www.ru.nl/fcdonders/fieldtrip/>) and custom MATLAB scripts. The continuous data were referenced to the average of all channels after interpolation and filtered with a high-pass, zero-phase Butterworth filter at 1 Hz. The data were then cut into epochs from 800 ms before to 2400 ms after speech onset, baseline correcting from  $-100$  to  $0$  ms relative to speech onset. Independent component analysis (ICA) was performed on the epoched data, constrained by the top 50 principle component analysis components. The topographic distributions of the top 20 ICA components discovered were screened for eye movement artifacts. No more than two components, with far frontal distributions clearly indicative of either eye blinks or lateral eye movements, were removed from the data for each subject. Epochs with shifts greater than  $\pm 80 \mu\text{V}$  were rejected.

Only trials associated with a correct behavioral response were included for EEG analysis. Because the Central Control condition was included only as a behavioral control, it was not included in any EEG analysis.

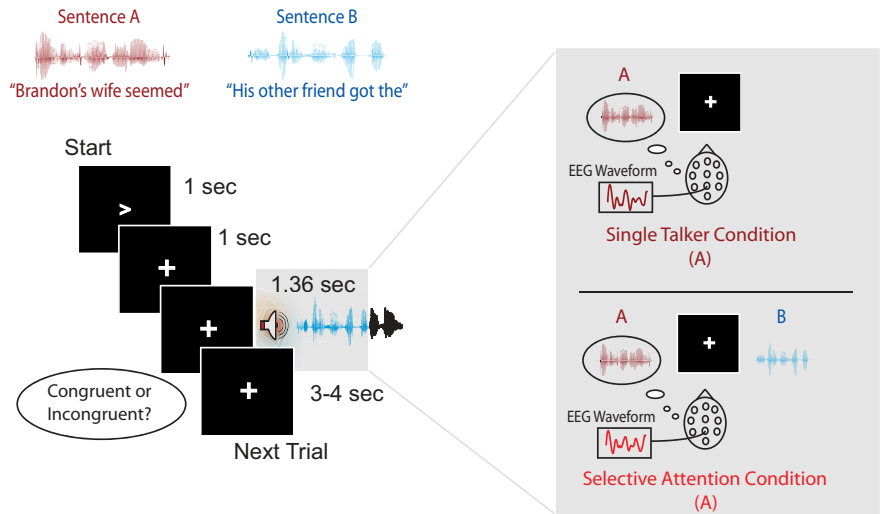
Epochs were sorted into separate bins for each participant based on the condition (Single Talker, Selective Attention) or on the content of the sentence (either sentence A or sentence B, as presented in the Single Talker condition or as attended in the Selective Attention conditions). The Selective Attention passive conditions (cue “0”) obviously could not be grouped based on attended content and were instead averaged only within presentation type. For some analyses, the Single Talker and Selective Attention conditions were further subdivided into bins based on the side of presentation (left or right) and direction of attention (left or right), respectively. For all bins, high- and low-pitch conditions were collapsed.

A two-equivalent-current dipole model based on the N1 component (100–150 ms) for the group grand averaged activity time locked to the onset of each sentence, collapsed across left and right Single Talker presentation, was created in Brain Electrical Source Analysis. The two dipoles were constrained to be symmetric in spatial location and allowed to fit freely to a single orientation for each dipole. The resulting fit placed vertically oriented dipoles in the left and right hemisphere with Talairach coordinates ( $x = 29$ ;  $y = -31$ ;  $z = 13$ ; for the right dipole) consistent with sources in or near Heschl’s gyrus. The residual variance for the fitting time interval was 2.9%. The location and orientation of the dipoles are consistent with the dipole parameters for sentence stimuli reported by Aiken and Picton (2008). In a subsequent step the dipole model was exported to MATLAB and was used as a spatial filter on the individual waveforms. The individual dipole waveforms were then filtered with eight different zero-phase bandpass Butterworth filters with frequencies of 1–4, 4–8, 8–12, 12–30, 30–50, 50–80, 80–120, and 120–160 Hz. Each filtered waveform was split into nine time windows, 341 ms long (1/4 of the incomplete sentence length), beginning with the window centered at the time of sentence onset (0 ms) and ending centered on sentence offset (1364 ms), shifting with 50% overlap. These waveforms, binned by participant, talker condition, sentence content, cue direction, side of presentation, hemisphere, frequency, and time, will subsequently be referred to as trial waveforms.

**Regression analysis.** To measure the frequency and time course of the ongoing representation of speech in auditory cortex, we developed a template-matching method. For each participant, an  $N$ -minus-1 group template waveform was created by averaging all trial waveforms not belonging to the current participant. There is thus no overlap in the data contained in the participant and group template waveforms, and bins within trial and group template waveforms can be collapsed independently. For all comparisons, the linear least-squares estimate between the template and all individual trial waveforms ( $\tau$ ) was calculated for each combination of comparison bins by the following equation:

$$\tau(T_{SC}, Y_{SC}) = \frac{\sum_{i=1}^n ((T'_{SC} T_{SC})^{-1} T'_{SC} Y_{SCi})}{n},$$

where  $T$  is the template waveform,  $Y$  represents individual trial waveforms, and  $S$  and  $C$  represent the sentence content and condition type, respectively. The  $\tau$  value, or regression coefficient, serves as an estimate of the extent to which the template waveform is present in each individual trial.  $\tau$  values for within- and across-sentence comparisons were found by the following equations:  $\tau_{\text{within}}(T_C, Y_C) = (\tau(T_{AC}, Y_{AC}) + \tau(T_{BC}, Y_{BC}))/2$  and  $\tau_{\text{across}}(T_C, Y_C) = (\tau(T_{AC}, Y_{BC}) + \tau(T_{BC}, Y_{AC}))/2$ , where  $A$  and  $B$  represent the presented/attended waveforms for sentences A and B, respectively. The discrimination index (DI) was calculated by simply sub-



**Figure 1.** Trial structure. A cue is presented to indicate the to-be-attended direction, followed by the onset of one of the sentences (Single Talker condition) or both simultaneously (Selective Attention condition). The expanded, inset box depicts a subject attending to sentence A while the EEG waveform is recorded. Upon hearing the final word, the participant then judges whether the attended sentence was congruent or incongruent.

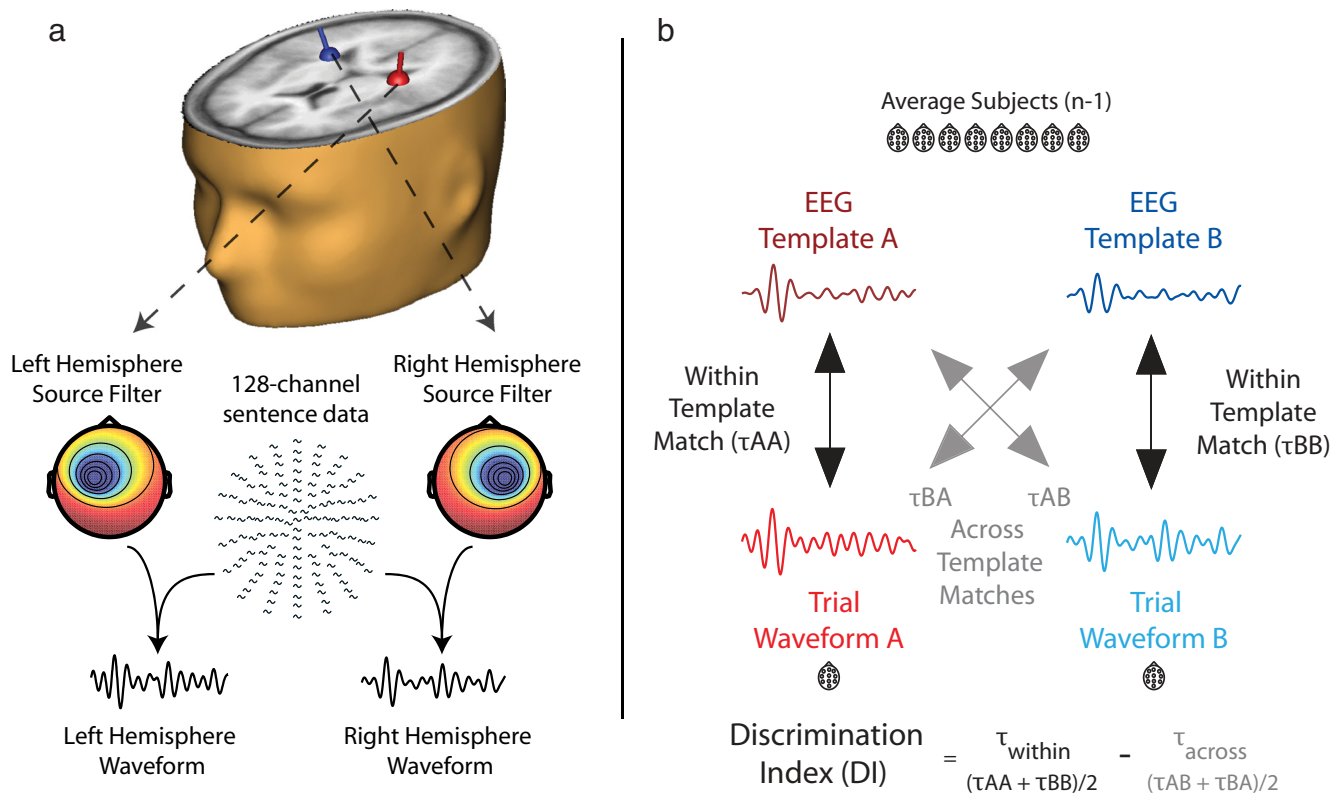
tracting the within and across  $\tau$  values as follows:  $DI(T_C, Y_C) = \tau_{\text{within}}(T_C, Y_C) - \tau_{\text{across}}(T_C, Y_C)$ .

A positive index reflects a shared signal between the trial and group EEG waveforms that distinguishes which of the two sentence waveforms was presented or attended on the individual trials (Fig. 2). A number of different comparisons between waveforms can be made, and the definition of the comparisons performed in this study can be found in Table 1. The method is conceptually similar to that devised by Luo and Poeppel (2007), but we choose to compare filtered waveforms rather than phase to be more sensitive to waveform phase changes within each analysis window and to quantify the magnitude and sign of enhancement and suppression.

For visualization across the scalp, topographic maps of the discrimination index are shown below (see Fig. 4) (also see supplemental Figure 1, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). The methods used were identical to those described above, with the exception that waveforms were derived from each channel of the 128-channel array, rather than the two source waveforms.  $T$ -scores at each channel were derived from a one-sample test of the individual mean DI across participants, with the null hypothesis of  $\mu_{DI} = 0$ .

For the calculation of enhancement and suppression in the Generalized Attentional Gain comparison, the within and across  $\tau$  values were subtracted by the responses in the Generalized Passive condition to remove any signal that could be attributed to stimulus attributes as follows:  $\tau_{\text{within-corrected}}(T_1, Y_2) = \tau_{\text{within}}(T_1, Y_2) - \tau(T_1, Y_2, p)$  and  $\tau_{\text{across-corrected}} = \tau_{\text{across}}(T_1, Y_2) - \tau(T_1, Y_2, p)$ , where the subscripts 1 and 2 represent active attention Single and Dual Talker conditions, respectively, and the subscript  $2P$  represents the Dual Talker Passive condition. The within and across  $\tau$  values now represent the enhancement and suppression of the attended signals, respectively, under the assumption that the time courses of  $A$  and  $B$  within any given analysis window are orthogonal. To remove any enhancement that could be an artifact of an increase in suppression, or vice versa, we applied the following final correction:  $\tau_{\text{enhance}}(T_1, Y_2) = \tau_{\text{within-corrected}}(T_1, Y_2) - \tau_{\text{across}}(T_1, Y_1)\tau_{\text{across-corrected}}(T_1, Y_2)$  and  $\tau_{\text{suppress}}(T_1, Y_2) = \tau_{\text{across-corrected}}(T_1, Y_2) - \tau_{\text{across}}(T_1, Y_1)\tau_{\text{within-corrected}}(T_1, Y_2)$ , where a positive  $\tau_{\text{enhance}}$  represents enhancement of the attended signal that cannot be attributed to suppression of the unattended signal and a negative  $\tau_{\text{suppress}}$  represents suppression of the unattended signal that cannot be attributed to enhancement of the attended signal.

**Alpha power analysis.** To measure changes in power in the alpha range, sensitive to the location of auditory spatial attention, the rms of the EEG signal across all 128 electrodes, filtered from 8 to 12 Hz, was measured for each of the nine time windows of the previous comparisons. The rms ampli-



**Figure 2.** EEG waveform analysis. *a*, Data from the 128-electrode array were reduced to two source waveforms in the left and right hemispheres, based on early-onset (N1) response to the first word of each sentence. Sources were localized in or near early auditory cortex. *b*, A diagram of the regression analysis used to produce the discrimination index. For a given subject, each trial EEG waveform is matched to a group average template for the same (“Within”) or a different (“Across”) sentence. A positive discrimination index means that the EEG signal can distinguish which sentence was presented or attended.

**Table 1. Discrimination index comparisons**

Comparison name	Group template waveform ( <i>T</i> ) condition	Trial waveform ( <i>Y</i> ) condition
Speech Encoding ( $T_1Y_1$ )	Single Talker	Single Talker
Attentional Gain ( $T_2Y_2$ )	Selective Attention	Selective Attention
Generalized Attentional Gain ( $T_1Y_2$ )	Single Talker	Selective Attention
Generalized Passive ( $T_1Y_{2p}$ )	Single Talker	Selective Attention (passive listening only)

tude of every trial was averaged, and the data were then collapsed across sentence content, leaving the two directions of attention (left and right) as conditions within the two main talker conditions (Single Talker and Selective Attention). The power for left-cued trials ( $P_{\alpha}(\text{CuedLeft})$ ) was then subtracted from right-cued trials ( $P_{\alpha}(\text{CuedRight})$ ) within each talker condition, to examine the differential response between the left and right sentence presentations in the Single Talker condition, and left attention versus right attention in the Selective Attention condition.

Based on previous visual studies showing alpha lateralization in posterior-parietal electrodes, we selected the 26 electrodes in the posterior left quadrant of the electrode array as the posterior left region of interest ( $ROI_{PL}$ ) and the 26 electrodes in the posterior right quadrant as the posterior right region of interest ( $ROI_{PR}$ ). We quantified this differential alpha power response across hemispheres to lateralized speech in a single measure: the alpha lateralization index (ALI), similar to the index of the same name described by Thut et al. (2006). We defined the alpha lateralization index with the following formula:  $ALI = ROI_{PL}(P_{\alpha}(\text{CuedLeft}) - P_{\alpha}(\text{CuedRight})) - ROI_{PR}(P_{\alpha}(\text{CuedLeft}) - P_{\alpha}(\text{CuedRight}))$ .

## Results

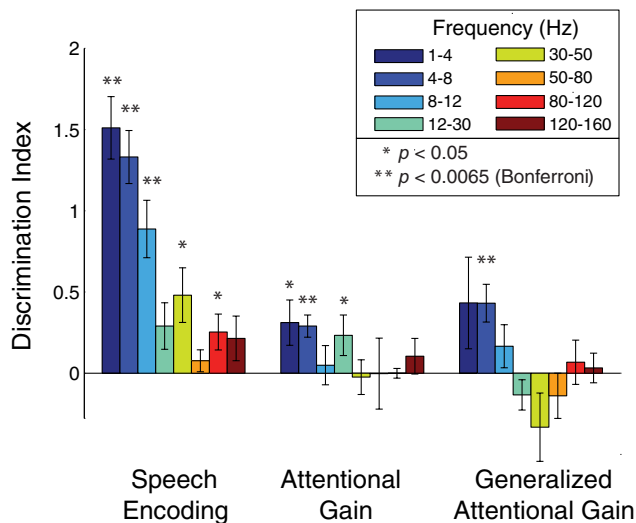
### Behavior

Participants performed the speech comprehension task with a high level of accuracy ( $M = 93.2\%$  correct,  $SD = 3.67\%$ ). A  $2 \times 2$

$2 \times 2$  ANOVA was performed on accuracy with the within-subject factors of talker condition (Single Talker, Selective Attention), direction of attention (left, right) and attended sentence content (sentence A, sentence B). A main effect of talker condition ( $F_{(1,12)} = 8.44, p = 0.0132$ ) was found, with significantly better accuracy for the Single Talker ( $M = 94.6\%$ ,  $SD = 3.63\%$ ) than the Selective Attention ( $M = 91.8\%$ ,  $SD = 4.44\%$ ) condition. A small but significant main effect of sentence content ( $F_{(1,12)} = 7.98, p = 0.0153$ ) was also found, with significantly better accuracy for the adjective sentence ( $M = 94.5\%$ ,  $SD = 3.33\%$ ) than the noun sentence ( $M = 91.9\%$ ,  $SD = 4.59\%$ ). A second ANOVA was performed based on the reaction time after the end of the sentence for correct trials, using the same factors. There was a main effect of sentence content ( $F_{(1,12)} = 12.2, p = 0.0044$ ) on reaction time, with the noun sentence ( $M = 1084$  ms,  $SD = 169$  ms) leading to slightly but significantly longer reaction times than the adjective sentence ( $M = 1028$  ms,  $SD = 153$  ms). There were no other significant main effects or interactions. Subject performance was therefore generally high and well balanced across stimuli in the Single Talker and Selective Attention Conditions. In a further Central Control condition, participant performance, though above chance (one-sided *t* test;  $t_{(12)} = 2.41, p = 0.016$ ), was extremely poor ( $M = 56.9\%$ ,  $SD = 10.2\%$ ), confirming that accurate performance of the Selective Attention task was dependent on spatial segregation.

### Frequencies of speech representation in auditory cortex consistent across individuals

We sought to identify which frequencies encode speech content consistently across individuals in auditory cortex. We therefore found the mean discrimination index value in the Speech Encod-



**Figure 3.** Gain control of speech representation across frequency. The DI is shown for three comparisons, across frequency. For Speech Encoding, a significant DI means that the EEG can distinguish sentence A from sentence B when presented alone. For Attentional Gain, a significant DI means that in a multitalker environment, attention modulates the EEG to distinguish which sentence was attended. In Generalized Attentional Gain, a significant DI means that to distinguish sentences, multitalker attention modulates the EEG signal that represents speech when presented alone. That is, attention changes the gain of the single-talker EEG signal.

ing comparison (Table 1), and collapsed across all dimensions except EEG frequency. A positive discrimination index means that the EEG signal at a particular frequency distinguishes between the sentences A and B when presented alone, in a consistent manner across subjects. Three of the eight frequencies had a discrimination index significantly above zero, based on one-tailed  $t$  tests, Bonferroni corrected for eight comparisons: 1–4 Hz ( $t_{(12)} = 7.54, p < 0.0001$ ), 4–8 Hz ( $t_{(12)} = 7.82, p < 0.0001$ ), and 8–12 Hz ( $t_{(12)} = 4.83, p = 0.0016$ ). An additional three frequencies had discrimination indices greater than zero with a  $p$  value  $< 0.05$ , without Bonferroni correction: 12–30 Hz ( $t_{(12)} = 1.91, p = 0.0373$ ), 30–50 Hz ( $t_{(12)} = 2.74, p = 0.0089$ ) and 80–120 Hz ( $t_{(12)} = 2.21, p = 0.024$ ) (Fig. 3). As expected, lower EEG frequencies were robust in discriminating which sentence was presented, even on individual trials. Unexpectedly, the 30–50 Hz and 80–120 Hz bands also showed positive discrimination indices, though not as strongly, which would suggest differential, phase-locked neural responses at very high frequencies that distinguished between sentences.

#### Frequencies of speech representation in auditory cortex modulated by selective attention

We then tested whether selective attention modulates these neural representations of speech. The same analysis steps were thus performed on the Attentional Gain comparison. Here, a positive discrimination index means that an individual's attention to a sentence causes the EEG signal to better match the group response when attending to that sentence, despite no difference in the stimulus presented. The Attentional Gain comparison had a significant discrimination index with Bonferroni correction at 4–8 Hz ( $t_{(12)} = 4.08, p = 0.0064$ ), with frequencies from 1 to 4 Hz ( $t_{(12)} = 2.15, p = 0.0266$ ) and 12–30 Hz ( $t_{(12)} = 1.80, p = 0.049$ ) having uncorrected  $p$  values  $< 0.05$ . Thus, selective attention causes a consistent, phase-locked response across individuals at lower frequencies, especially at 4–8 Hz, that distinguishes which sentence is being attended.

Notice, this comparison is indifferent to whether the speech EEG waveform for one talker presented alone generalizes to, or is qualitatively similar in, multitalker situations. Rather, it only requires that attention consistently modulates the EEG signal when listeners selectively attend in the presence of multiple talkers. To test whether selective attention during competing speech acts via gain control on the neural responses that represent speech when heard alone, we performed the same analysis steps on the Generalized Attentional Gain comparison. Here, a positive discrimination index means that attending to a sentence causes an individual's EEG signal to better match the group signal for that sentence presented alone, despite no differences in the stimuli presented. The Generalized Attentional Gain comparison had a significant discrimination index only at 4–8 Hz ( $t_{(12)} = 3.57, p = 0.015$ , corrected), with no other significant frequency ranges. Importantly, this attentional gain could not be explained by a simple enhancement of the traditional early-onset (N1) response to words, as shown by modeling sentence responses as a series of word onset transients (results available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). Nor can these results be explained by increased phase entrainment of an intrinsic 4–8 Hz oscillation (also available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). Thus, attending to a speech signal increased its continuous neural representation compared with the unattended sentence, in the 4–8 Hz range.

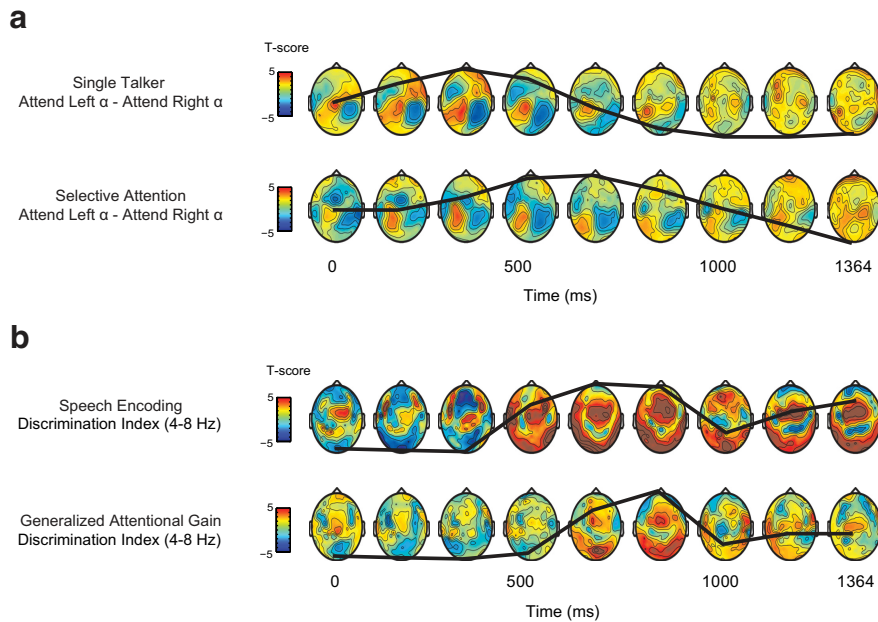
#### Hemispheric differences in speech representation

Further analysis of source hemisphere was performed only on the frequency ranges with a significant discrimination index in the previous analysis. Separate paired two sample  $t$  tests were performed on the 1–4 Hz, 4–8 Hz and 8–12 Hz waveforms from the Speech Encoding comparison. There was a significant main effect of source hemisphere only for the 4–8 Hz range ( $t_{(12)} = 3.40, p = 0.0053$ ), with the right hemisphere having a significantly greater discrimination index than the left hemisphere.

The same analysis was performed with the Attentional Gain comparison at the 4–8 Hz range, again finding greater discrimination in the right versus left source ( $t_{(12)} = 2.31, p = 0.040$ ). In the Generalized Attentional Gain condition, the 4–8 Hz frequency range was compared, and though again the right hemisphere appeared to be more discriminative than the left hemisphere, this difference was not quite significant ( $t_{(12)} = 2.13, p = 0.054$ ). In general, the cortical response in the right hemisphere was more robust in predicting which sentences were presented or attended, in the frequency range of 4–8 Hz.

#### Enhancement of the speech representation by selective attention

While the discrimination index distinguishes EEG waveforms produced by attending to a particular sentence while ignoring the other, it cannot show whether discrimination results from enhancement of the attended sentence signal and/or suppression of the unattended sentence signal. To test each possibility, attended and unattended activity was compared with a passive listening “baseline.” Specifically, the passive listening template-matching regression coefficients were subtracted from the attended and unattended regression coefficients, forming an index of enhancement and suppression, respectively. This was performed for all participants with positive selective attention discrimination indices at 4–8 Hz ( $n = 10$ ). These values were further corrected to discount any values resulting from nonorthogonality in the two template waveforms (see Materials and Methods). A positive value in this index means the attended signal was enhanced, while



**Figure 4.** Spatial attention and sentence content selection. *a*, Scalp topographic maps over time of alpha activity following the onset of the sentences (left minus right cued conditions). The superimposed line plots represent the normalized time course of the Single Talker and Selective Attention alpha lateralization indices. *b*, Topographic maps of the discrimination indices following the onset of the sentences. The superimposed line plots represent the normalized time course of the Speech Encoding and Generalized Attentional Gain discrimination indices, averaged from the two source waveforms.

a negative value indicates that the unattended signal was suppressed.

At 4–8 Hz, the index on enhancement of the attended speech was significantly greater than zero, as revealed with a one-tailed  $t$  test ( $t_{(9)} = 1.90, p = 0.045$ ), while the index of suppression of the unattended speech was not significantly different from zero, showing a weak negative trend in a one-tailed test ( $t_{(9)} = -0.96, p = 0.18$ ). Auditory selective attention to continuous speech therefore acts at least via an enhancement of the attended signal, as opposed to a strong suppression of the unattended signal.

### Auditory spatial attention results in differential hemispheric alpha power

Having indexed the content of selective attention to speech, we tested whether alpha power in posterior cortex could predict the location of speech presentation and selective attention. Based on studies in the visual domain, we expected relative contralateral alpha suppression and ipsilateral alpha enhancement when attention is focused laterally. As can be seen in Figure 4*a*, presenting a sentence to the left versus right in virtual auditory space (Single Talker condition) induces a clear difference between the hemispheres at parietal channels due to stimulus location, with relative ipsilateral alpha enhancement and relative contralateral alpha suppression. We performed a Student's  $t$  test using the alpha lateralization index for the Single Talker condition ( $ALI_{ST}$ ) across all participants and found it was significantly greater than zero ( $t_{(12)} = 3.14, p = 0.0043$ ), meaning that lateralized alpha power was significantly predictive of the side of speech presentation. When both sentences are presented simultaneously (Selective Attention condition), the differential alpha response based solely on selective attention appears to have a similar time course and topography as the Single Talker condition (Fig. 4*a*). A Student's  $t$  test of the Selective Attention alpha lateralization index ( $ALI_{SA}$ ) revealed the index was significantly greater than zero ( $t_{(12)} = 2.37, p = 0.018$ ), meaning lateralized alpha power predicted the

direction of auditory selective attention in the absence of stimulus differences, in a manner similar that of visual-spatial selective attention.

### Signals of speech representation and selection over time

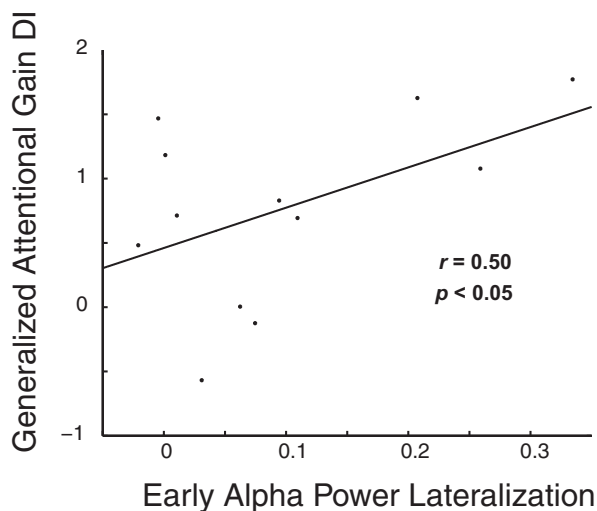
The neural mechanisms of allocating spatial attention and selecting a talker may not occur uniformly through time. We therefore examined the time course of both alpha lateralization and the discrimination index over the duration of the sentence, shown normalized in magnitude in Figure 4, *a* and *b*. A one-way ANOVA of the  $ALI_{ST}$  across time (9 time windows), collapsed over all other bins, revealed a significant main effect of time ( $F = 7.34, p < 0.0001$ ), starting at a value significantly above zero and peaking at  $\sim 340$  ms. There was also a significant main effect of time for the  $ALI_{SA}$  ( $F = 2.9, p < 0.0057$ ), though peaking somewhat later at 682 ms.

A one-way ANOVA of the Speech Encoding discrimination index at 4–8 Hz across time (9 time windows), collapsed over all other bins, also revealed a significant main effect of time ( $F = 18.42, p < 0.0001$ ) as did the Generalized Attentional Gain discrimination index ( $F = 3.24, p = 0.0026$ ). The time course of the Generalized Attentional Gain index appears to be quite similar to that of the Speech Encoding index. This is not surprising, as the ability to discriminate sounds based on selective attention should depend on the degree to which their responses can be separated when presented alone. However, as with the alpha lateralization index, the signal due purely to selective attention (Generalized Attentional Gain DI) peaks later than the signal evoked by external stimulus differences (Speech Encoding DI), suggesting there may be a buildup period for gain by selective attention. Consistent with this, the spatial alpha effect tends to dominate early in the sentence, while the attentional discrimination effect dominates later in the sentence.

### Lateralized alpha power and attentional gain

Given that both phase-locked sentence-specific responses at 4–8 Hz and non-phase-locked power at 8–12 Hz are sensitive to attention among multiple talkers, we sought to test whether there is a relationship between the two signals. We predicted that in the Selective Attention condition, participants with greater early alpha lateralization, which is linked to attending to a particular location, would also tend to have a greater Generalized Attentional Gain discrimination index, which is a measure of selecting the sentence content at the attended location. To test this, we calculated the Pearson's correlation coefficient between the mean Selective Attention alpha lateralization index early in the sentence (0–512 ms) and the mean Generalized Attentional Gain discrimination index later in the sentence (682–1364 ms) for each participant (Fig. 5). A significant, positive correlation was found between the two indices ( $r = 0.495, t_{(12)} = 1.89, p = 0.043$ ).

Although the topographies of these signals were quite different, it is possible that this correlation reflected individual differences in factors independent of selective attention. However, we found no significant correlation between the early Speech Encod-



**Figure 5.** Early alpha lateralization predicts generalized attentional gain. A scatter plot of the average early Selective Attention alpha lateralization index versus late Generalized Attentional Gain discrimination index for each individual participant, with the line of best fit. Parietal alpha lateralization predicts how strongly attention modulates the selected speech representation.

ing DI and late  $ALI_{SA}$  ( $r = -0.148$ ), ruling out general differences in an individual's stimulus driven response as the cause of the relationship. We could also rule out that individual differences in alpha activity over the entire time course caused this relationship, as there was no significant correlation when the  $ALI_{SA}$  was taken from the same time window as the late Generalized DI ( $r = 0.074$ ), nor if the  $ALI_{SA}$  was replaced with the overall alpha power in the same time period ( $r = 0.17$ ). Thus, the relationship between lateralized alpha activity and the gain of sentence content reflects an interaction between distinct mechanisms of spatial and selective attention.

## Discussion

We have shown that selective attention in a “cocktail party” modulates the early cortical representation of speech via a gain mechanism. Specifically, selective attention increases discrimination of the attended speech signal in auditory cortex in the range of 4–8 Hz, a frequency band strongly represented in the speech envelope and known to be important for speech comprehension. We demonstrate furthermore that this attentional gain is due to enhancement of the attended sentence, and possibly suppression of the unattended stream.

In addition to the neural gain in speech representation, our results establish that alpha power lateralization at parieto-occipital sites reflects the direction of auditory attention to continuous speech in space. Posterior parietal cortical involvement in attentional selection of speech-in-noise has been shown with high spatial resolution using fMRI, with greater bilateral superior parietal lobule (SPL) activity when participants select a talker based on spatial attributes rather than pitch (Hill and Miller, 2009) and for shifts versus maintenance of auditory selective attention (Shomstein and Yantis, 2006). Furthermore, a recent MEG study found occipitoparietal alpha activity when subjects maintained lateralized sounds in working memory (Kaiser et al., 2007), and recent fMRI studies find sensitivity to auditory spatial attention in occipital as well as parietal cortex (Wu et al., 2007; Cate et al., 2009). Notably, the topography of our alpha lateralization is nearly identical to that in cued visuospatial attention (Worden et al., 2000; Sauseng et al., 2005; Kelly et al., 2006; Rihs

et al., 2009) and intermodal attentional switching (Foxe et al., 1998). Although the arrow cue in the current experiment could have evoked pure visuospatial attention with alpha lateralization, this interpretation is unlikely as the cue onset occurred long before (2 s) the ALI analysis window and no visual stimuli were ever colocalized with the voices. The overlap of alpha modulation at parieto-occipital sites for both auditory and visuospatial attention adds to growing behavioral and physiological evidence for a supramodal mechanism of attentional selection (Farah et al., 1989; Spence and Driver, 1996; Eimer et al., 2004).

Not only does alpha power at parieto-occipital sites reflect where the brain allocates selective attention to continuous speech in space, but it also predicts how well the auditory cortex distinguishes which sentence is attended. The correlation between alpha power lateralization and the strength of the selective attention response to continuous speech provides a mechanistic link between parieto-occipital alpha activity and selective enhancement of the attended auditory stimulus. The time courses of our chosen measures of selective attention also suggest an order to these effects. Alpha lateralization due to spatial attention peaks early and disappears before the end of the sentence, which implies that differential parieto-occipital activity is needed to select an auditory object in space, but is not required to maintain the auditory stream over time. This is consistent with fMRI evidence for greater activity in the SPL to auditory speech stream switching than maintenance (Shomstein and Yantis, 2006). The time course of the neural selection of speech content, as indexed by the Generalized Attentional Gain comparison, peaks later and is sustained throughout the sentence. This time course is more difficult to interpret, and could be well explained by differences in the stimulus envelope over time, such that the more the envelopes of two sentences differ, the easier it is to detect a difference in the neural response between them (Abrams et al., 2008; Aiken and Picton, 2008). However, other cognitive temporal effects, such as perceptual buildup in the streaming of the attended sentences (Shinn-Cunningham and Best, 2008), or the increased task relevance around sentence endings in our paradigm, may also play a role. Future experiments with a variety of longer and systematically varied sentence structures are needed to disambiguate the issue. Regardless, the attentional gain of sentence content clearly continues after the offset of alpha lateralization, implying that alpha spatial selection mechanism is resolved before the entire sentence has been processed.

Though the current study shows clear attentional enhancement of a 4–8 Hz signal in auditory cortex, our approach is agnostic with respect to the representational nature of the signal itself. As in similar, previous studies (Suppes et al., 1998, 1999; Luo and Poeppel, 2007), we find robust speech representations in these lower frequencies. Further analysis (results available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material) established that the low-frequency attentional gain is an ongoing phenomenon, as with ongoing attention to tone sequences (Elhilali et al., 2009), and cannot be explained by a traditional, transient word onset N1 response. Several possibilities for the nature of this ongoing signal remain. Most likely, much of the initial 4–8 Hz signal reflects a response to the speech envelope, which has substantial power in the 2–20 Hz range (Purcell et al., 2004; Aiken and Picton, 2008) and is known to evoke a following response preferentially at the natural frequencies of speech envelope (Ahissar et al., 2001). Furthermore, this low-frequency encoding is greater in the right than left hemisphere, also consistent with prior work on the speech envelope (Tremblay and Kraus, 2002; Abrams et al., 2008), as well as asymmetric sampling in time theory, in which syllabic timescales

are processed preferentially in the right hemisphere (Poeppl, 2003; Giraud et al., 2007; Abrams et al., 2008; Overath et al., 2008). An alternative view is that the 4–8 Hz signal is not a speech representation per se but rather reflects intrinsic oscillatory neural activity that is phase reset by an ongoing stimulus (Makeig et al., 2002; Luo and Poeppl, 2007; Lakatos et al., 2008; Bonte et al., 2009). Although possible, in line with recent papers in the visual domain (Mazaheri and Jensen, 2006; Risner et al., 2009), we found little evidence that the 4–8 Hz signal from auditory cortex before the stimulus maintained phase information for multiple cycles (results available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). A parsimonious interpretation of our results suggests that this speech encoding signal is closely related to the properties of stimulus acoustics, such as the speech envelope, and limited by the temporal resolution of auditory cortical networks.

While previous studies on the cortical response frequencies to natural continuous speech report no differentiating frequencies >50 Hz (Suppes et al., 1998, 1999; Ahissar et al., 2001; Bidet-Caullet et al., 2007; Luo and Poeppl, 2007; Buiatti et al., 2009), we found weak evidence of neural speech encoding >80 Hz. At 80–120 Hz, the finding of a positive discrimination index requires a phase-locked response, shared among participants with latency differences of less than ~6 ms (1/2 the longest wavelength). Higher frequencies may reflect transient responses to particular plosives, fricatives, or transitions within the speech stream or possibly higher order processes, such as matching external auditory stimuli to templates in working memory (Kaiser et al., 2003, 2008; Lenz et al., 2008; Shahin et al., 2009). Further studies are required to determine whether phase-locked speech encoding responses are truly present in cortex at these frequencies.

Our attentional gain analyses were performed using a dipole model aimed at maximizing signal originating in or near early auditory cortex. Such generators produce a canonical fronto-central and posterior scalp distribution and can be modeled to a large extent by one dipole in each hemisphere (Scherg et al., 2007; Aiken and Picton, 2008). We expected that most time-locked cortical activity in response to an extended, complex acoustic stimulus such as speech would come from this region of cortex, producing a similar pattern on the scalp. Indeed, the discrimination index at 4–8 Hz across all scalp sites is largely consistent with our two dipole source model based on the early auditory onset response (N1) (see supplemental Fig. 1, available at [www.jneurosci.org](http://www.jneurosci.org) as supplemental material). This finding agrees with the MEG results of Luo and Poeppl, which found sentence discrimination greatest in the same channels as the auditory M100 (Luo and Poeppl, 2007). Nevertheless, this approach cannot exclude the possibility that other regions, particularly language related areas such as superior temporal sulcus, may also have time-locked, differential responses between sentences that contribute to the source waveforms.

We should emphasize that many of our results rely on a novel application of EEG template matching to produce the discrimination index. This technique has several advantages when compared with traditional ERP or oscillatory power analysis. Notably, it allows for the measurement of a shared response pattern to an extended stimulus without requiring any prior knowledge of the pattern, and with very few assumptions about its time course and frequency content. This makes the method ideal for exploratory studies in which, unlike traditional ERP analysis, unexpected but consistently time-locked responses can be detected. It has some limitations, including the requirement that the same stimulus must be presented multiple times and that the responses must be phase locked. But as a complement to more established

methods, it offers substantial advantage in characterizing the multiple, temporally overlapping signals so common in naturalistic environments.

We propose that auditory selective attention in a cluttered, realistic environment begins with allocating spatial attention. This is evidenced by increased contralateral alpha suppression, which may reflect networks in the contralateral posterior parietal cortex shifting from a passive to active state, or active suppression of the unattended space. The posterior parietal cortex may then bind the auditory object to a location in space to assist in the selection and streaming of the auditory content. Once the object is successfully streaming, supramodal spatial activity reduces and stream selection continues based on nonspatial as well as spatial cues. In contrast to the suppression of non-phase-locked alpha in the parieto-occipital cortex, phase-locked early auditory cortical representation of the attended stream in the temporal lobe is then enhanced via a gain mechanism, leading to successful comprehension.

## References

- Abrams DA, Nicol T, Zecker S, Kraus N (2008) Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J Neurosci* 28:3958–3965.
- Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci U S A* 98:13367–13372.
- Aiken SJ, Picton TW (2008) Human cortical responses to the speech envelope. *Ear Hear* 29:139–157.
- Alcaini M, Giard MH, Echallier JF, Pernier J (1995) Selective auditory attention effects in tonotopically organized cortical areas: a topographic ERP study. *Hum Brain Mapp* 2:159–169.
- Bidet-Caullet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O (2007) Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J Neurosci* 27:9252–9261.
- Bonte M, Valente G, Formisano E (2009) Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. *J Neurosci* 29:1699–1706.
- Broadbent DE (1957) A mechanical model for human attention and immediate memory. *Psychol Rev* 64:205–215.
- Buiatti M, Peña M, Dehaene-Lambertz G (2009) Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage* 44:509–519.
- 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.
- Coch D, Sanders LD, Neville HJ (2005) An event-related potential study of selective auditory attention in children and adults. *J Cogn Neurosci* 17:605–622.
- Dajani HR, Purcell D, Wong W, Kunov H, Picton TW (2005) Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Trans Biomed Eng* 52:1614–1618.
- Driver J (2001) A selective review of selective attention research from the past century. *Br J Psychol* 92:53–78.
- Eimer M, van Velzen J, Driver J (2004) ERP evidence for cross-modal audiovisual effects of endogenous spatial attention within hemifields. *J Cogn Neurosci* 16:272–288.
- Elhilali M, Xiang J, Shamma SA, Simon JZ (2009) Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biol* 7:e1000129.
- Farah MJ, Wong AB, Monheit MA, Morrow LA (1989) Parietal lobe mechanisms of spatial attention: modality-specific or supramodal? *Neuropsychologia* 27:461–470.
- Foxe JJ, Simpson GV, Ahlfors SP (1998) Parieto-occipital ~10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport* 9:3929–3933.
- Giraud AL, Kleinschmidt A, Poeppl D, Lund TE, Frackowiak RS, Laufs H



- (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* 56:1127–1134.
- Gruber WR, Klimesch W, Sauseng P, Doppelmayr M (2005) Alpha phase synchronization predicts P1 and N1 latency and amplitude size. *Cereb Cortex* 15:371–377.
- Hill KT, Miller LM (2009) Auditory attentional control and selection during cocktail party listening. *Cereb Cortex*. Advance online publication. Retrieved Dec. 31, 2009. doi:10.1093/cercor/bhp124.
- Kaiser J, Ripper B, Birbaumer N, Lutzenberger W (2003) Dynamics of gamma-band activity in human magnetoencephalogram during auditory pattern working memory. *Neuroimage* 20:816–827.
- Kaiser J, Heidegger T, Wibral M, Altmann CF, Lutzenberger W (2007) Alpha synchronization during auditory spatial short-term memory. *Neuroreport* 18:1129–1132.
- Kaiser J, Heidegger T, Wibral M, Altmann CF, Lutzenberger W (2008) Distinct gamma-band components reflect the short-term memory maintenance of different sound lateralization angles. *Cereb Cortex* 18:2286–2295.
- Kelly SP, Lalor EC, Reilly RB, Foxe JJ (2006) Increases in alpha oscillatory power reflect an active retinotopic mechanism for distracter suppression during sustained visuospatial attention. *J Neurophysiol* 95:3844–3851.
- Krishnan A (2002) Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 166:192–201.
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320:110–113.
- Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J Neurophysiol* 102:349–359.
- Langendijk EH, Bronkhorst AW (2000) Fidelity of three-dimensional-sound reproduction using a virtual auditory display. *J Acoust Soc Am* 107:528–537.
- Lavie N (1995) Perceptual load as a necessary condition for selective attention. *J Exp Psychol Hum Percept Perform* 21:451–468.
- Lenz D, Jeschke M, Schadow J, Naue N, Ohl FW, Herrmann CS (2008) Human EEG very high frequency oscillations reflect the number of matches with a template in auditory short-term memory. *Brain Res* 1220:81–92.
- Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001–1010.
- Makeig S, Westerfield M, Jung TP, Enghoff S, Townsend J, Courchesne E, Sejnowski TJ (2002) Dynamic brain sources of visual evoked responses. *Science* 295:690–694.
- Mazaheri A, Jensen O (2006) Posterior  $\alpha$  activity is not phase-reset by visual stimuli. *Proc Natl Acad Sci U S A* 103:2948–2952.
- Medendorp WP, Kramer GF, Jensen O, Oostenveld R, Schoffelen JM, Fries P (2007) Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. *Cereb Cortex* 17:2364–2374.
- Musacchia G, Sams M, Skoe E, Kraus N (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci U S A* 104:15894–15898.
- Näätänen R, Michie PT (1979) Early selective-attention effects on the evoked potential: a critical review and reinterpretation. *Biol Psychol* 8:81–136.
- Nager W, Dethlefsen C, Münte TF (2008) Attention to human speakers in a virtual auditory environment: brain potential evidence. *Brain Res* 1220:164–170.
- Overath T, Kumar S, von Kriegstein K, Griffiths TD (2008) Encoding of spectral correlation over time in auditory cortex. *J Neurosci* 28:13268–13273.
- Palva S, Palva JM (2007) New vistas for  $\alpha$ -frequency band oscillations. *Trends Neurosci* 30:150–158.
- Pichora-Fuller MK, Singh G (2006) Effects of age on auditory and cognitive processing: implications for hearing aid fitting and audiologic rehabilitation. *Trends Amplif* 10:29–59.
- Picton TW, Hillyard SA (1974) Human auditory evoked potentials. II: Effects of attention. *Electroencephalogr Clin Neurophysiol* 36:191–199.
- Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time.' *Speech Commun* 41:245–255.
- Purcell DW, John SM, Schneider BA, Picton TW (2004) Human temporal auditory acuity as assessed by envelope following responses. *J Acoust Soc Am* 116:3581–3593.
- Rihs TA, Michel CM, Thut G (2009) A bias for posterior  $\alpha$ -band power suppression versus enhancement during shifting versus maintenance of spatial attention. *Neuroimage* 44:190–199.
- Risner ML, Aura CJ, Black JE, Gawne TJ (2009) The visual evoked potential is independent of surface alpha rhythm phase. *Neuroimage* 45:463–469.
- 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.
- Scherg M, Vajsar J, Picton TW (2007) A source analysis of the late human auditory evoked potentials. *J Cogn Neurosci* 1:336–355.
- Senkowski D, Saint-Amour D, Gruber T, Foxe JJ (2008) Look who's talking: the deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions. *Neuroimage* 43:379–387.
- Shahin AJ, Picton TW, Miller LM (2009) Brain oscillations during semantic evaluation of speech. *Brain Cogn* 70:259–266.
- Shinn-Cunningham BG, Best V (2008) Selective attention in normal and impaired hearing. *Trends Amplif* 12:283–299.
- Shomstein S, Yantis S (2006) Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *J Neurosci* 26:435–439.
- Spence C, Driver J (1996) Audiovisual links in endogenous covert spatial attention. *J Exp Psychol Hum Percept Perform* 22:1005–1030.
- Suppes P, Lu ZL, Han B (1997) Brain wave recognition of words. *Proc Natl Acad Sci U S A* 94:14965–14969.
- Suppes P, Han B, Lu ZL (1998) Brain-wave recognition of sentences. *Proc Natl Acad Sci U S A* 95:15861–15866.
- Suppes P, Han B, Epelboim J, Lu ZL (1999) Invariance between subjects of brain wave representations of language. *Proc Natl Acad Sci U S A* 96:12953–12958.
- Teder W, Kujala T, Näätänen R (1993) Selection of speech messages in free-field listening. *Neuroreport* 5:307–309.
- Thut G, Nietzel A, Brandt SA, Pascual-Leone A (2006)  $\alpha$ -Band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J Neurosci* 26:9494–9502.
- Treisman A, Geffen G (1967) Selective attention: perception or response? *Q J Exp Psychol* 19:1–17.
- Tremblay KL, Kraus N (2002) Auditory training induces asymmetrical changes in cortical neural activity. *J Speech Lang Hear Res* 45:564–572.
- Woods DL, Alho K, Algazi A (1993) Intermodal selective attention: evidence for processing in tonotopic auditory fields. *Psychophysiology* 30:287–295.
- Worden MS, Foxe JJ, Wang N, Simpson GV (2000) Anticipatory biasing of visuospatial attention indexed by retinotopically specific  $\alpha$ -band electroencephalography increases over occipital cortex. *J Neurosci* 20:RC63.
- Wu CT, Weissman DH, Roberts KC, Woldorff MG (2007) The neural circuitry underlying the executive control of auditory spatial attention. *Brain Res* 1134:187–198.