Hide Cover

[Research Articles]

Human Cortical Responses to the Speech Envelope

Aiken, Steven J.; Picton, Terence W.

Author Information

Rotman Research Institute, Baycrest Centre for Geriatric Care, University of Toronto, Toronto, Canada.
Author's Note: Steven J. Aiken is now at the School of Human Communication Disorders, Dalhousie University, Halifax, Canada.
Address for correspondence: Steven J Aiken, School of Human Communication Disorders, Dalhousie University, 5599 Fenwick Street, Halifax, Nova Scotia B3H1R2, Canada. E-mail: steve.aiken@dal.ca.
Received March 9, 2007; accepted August 8, 2007.

Abstract

Objective: To evaluate the response of the human auditory cortex to the temporal amplitude-envelope of speech. Responses to the speech envelope could be useful for validating the neural encoding of intelligible speech, particularly during hearing aid fittings—because hearing aid gain and compression characteristics for ongoing speech should more closely resemble real world performance than for isolated brief syllables.

Design: The speech envelope comprises energy changes corresponding to phonemic and syllabic transitions. Envelope frequencies between 2 and 20 Hz are important for speech intelligibility. Human event-related potentials were recorded to six different sentences and the sources of these potentials in the auditory cortex were determined. To improve the signal to noise ratio over ongoing electroencephalographic recordings, we averaged the responses over multiple presentations, and derived source waveforms from multichannel scalp recordings. Source analysis led to bilateral, symmetrical, vertical, and horizontal dipoles in the posterior auditory cortices. The source waveforms were then cross-correlated with the low frequency log-envelopes of the sentences. The significance and latency of the maximum correlation for each sentence demonstrated the presence and latency of the brain's response. The source waveforms were also cross-correlated with a simple model based on a series of overlapping transient responses to stimulus change (the derivative of the log-envelope).

Results: Correlations between the log-envelope and vertical dipole source waveforms were significant for all sentences and for all but one of the participants (mean $r = 0.35$), at an average delay of 175 (left) to 180 (right) msec. Correlations between the transient response model (P1 at 68 msec, N1 at 124 msec, and P2 at 208 msec) and the vertical dipole source waveforms were detected for all sentences and all participants (mean $r = 0.30$), at an average delay of 6 (right) to 10 (left) msec.

Conclusions: These results show that the human auditory cortex either directly follows the speech

envelope or consistently reacts to changes in this envelope. The delay between the envelope and the response is approximately 180 msec.

Introduction

The human brain responds differently to speech than to other sounds. Speech stimuli elicit different patterns of hemodynamic activity than nonspeech sounds. Generally there is greater activation of the left hemisphere (Price, et al., 2005; Tervaniemi & Hugdahl, 2003), although the degree of hemispheric asymmetry varies greatly with stimulus variations and task demands (Josse & Tzourio-Mazoyer, 2004). Speech stimuli also elicit different patterns of electrical and magnetic activity than acoustically similar nonspeech sounds (Eulitz, et al., 1995; Mäkelä, et al., 2003; Shtyrov, et al., 2005; Tiitinen, et al., 1999), particularly when the speech sounds are presented in a linguistic context (Shtyrov, et al., 2005). A complete understanding of the auditory system may therefore require studies of responses to both nonspeech and speech stimuli.

Electrical and magnetic responses are particularly well suited for studies using speech stimuli, because their temporal resolution is sufficient to reflect the rapid processing necessary for speech perception. Electrical responses to speech are more likely to be abnormal than responses to nonspeech stimuli in patients with learning disabilities (Purdy, et al., 2002), dyslexia (Alonso-Bua, et al., 2006; Schulte-Korne, et al., 1998), specific language impairment (Uwer, et al., 2002), and aphasia (Ilvonen, et al., 2004). Electrical responses to speech may also be useful for validating hearing aid fittings in individuals who cannot provide behavioral responses (e.g., infants), because nonspeech stimuli can alter the gain and compression characteristics of a hearing aid in ways that are difficult to predict. Acoustic real ear measurements are generally used to verify that a hearing aid is providing appropriate gain at the frequencies necessary for speech. However, adjusting a hearing aid to bring speech sounds to an intensity that can be heard does not ensure that the brain will discriminate between different speech sounds—audibility does not necessarily entail intelligibility. Evoked potentials to speech presented at suprathreshold levels are relatively easy to record, and might be used to validate that intelligible speech has been registered in the nervous system.

The information in speech is organized at many levels from the pitch of the glottal source to the rhythms of prosody. Evoked potentials can be recorded to changes at each of these levels. Voiced speech segments are characterized by the glottal source frequency and its harmonics, which are spectrally shaped by resonances (formants) created by the vocal tract. Envelope and frequency-following responses can be evoked by the glottal source frequency (Aiken & Picton, 2006; Krishnan, et al., 2004; Russo, et al., 2004) and its harmonics (Krishnan, 2002; Plyler & Krishnan, 2001). These responses are mainly generated in the brainstem (Herdman, et al., 2002; Smith, et al., 1975).

A limitation of this approach is that envelope- and frequency-following responses can only create a temporal representation of sound when neurons become synchronized or phase-locked to a stimulus or its envelope. However, much useful information in speech is at frequencies above 1500 Hz (e.g., Studebaker & Sherbecoe, 2002), where phase-locking is unlikely to occur. This problem could be approached by recording responses to the slower changes in frequency or amplitude that characterize speech stimuli. Dimitrijevic et al. (2004) reported relations between word recognition performance and the presence of multiple steady state responses to carrier frequencies within the expected intensity-frequency range of human speech. Unfortunately, because many hearing aids are designed to vary gain characteristics on the basis of ongoing modulations present in the input (e.g., to attenuate background noise), modulating or acoustically modifying speech could cause a hearing aid to process the speech differently.

Another solution is to record cortical transient responses to high frequency speech elements, such as plosive bursts and frication onsets and offsets (e.g., Agung, et al., 2006; Kelly, et al., 2005; Ostroff, et al., 1998; Tremblay, et al., 2006a,b). Cortical transient responses to various acoustic changes in speech can be measured reliably (Martin, et al., 2007; Tremblay, et al., 2003a) and are sensitive to audibility (Korczak, et al., 2005; Martin, et al., 1997; Martin & Boothroyd, 1999; Oates, et al., 2002; Tremblay, et al., 2003a) as well as to other factors such as the subject's age (Tremblay, et al., 2002, 2003b, 2004) and the integrity of the auditory system (Kraus, et al., 2000; Rance, et al., 2002). A reliable transient response to a particular speech element, such as a plosive burst, would indicate that the auditory system has registered the event in some fashion (Eggermont & Ponton, 2002), thereby validating the hearing aid fitting for that particular speech sound. However, speech typically involves a large number of simultaneous changes. For instance, the transition from a voiced plosive consonant to a diphthong involves a silent period followed by the onset of voicing, and a plosive burst with simultaneous formant movement. Transient responses could occur in response to each of these changes, and might not be easy to distinguish in the scalp response. This complexity is typically reduced by recording a large number of responses to simple stimuli that involve a small number of acoustic changes [e.g., consonant-vowel (CV) syllables]. Unfortunately, these brief stimuli have amplitude envelopes that differ considerably from continuous speech, and might be treated differently from continuous speech by nonlinear hearing aids (Scollie & Seewald, 2002; Stone & Moore, 1992). In particular, short CV syllables may be subjected to less compression than continuous speech, which could lead to an overestimation of hearing aid gain and output, especially in hearing aids with long compression-release time constants. This problem is augmented by the ubiquity of digital signal processing algorithms that vary hearing aid gain and compression parameters on the basis of signal envelope characteristics.

Instead of recording transient responses to individual acoustic changes, it should be possible to record an envelope-following response to the slow temporal amplitude envelope of longer speech segments such as sentences. Demonstrating these responses could show that envelope information has been registered in the auditory system. The hearing aid gain and compression characteristics for these longer lasting stimuli should more closely resemble real world performance than for isolated brief syllables.

The speech envelope comprises energy changes corresponding to phonemic and syllabic transitions, and is most intense below 10 Hz. Envelope-following responses have been successfully recorded at similar rates using sinusoidally modulated white noise (Liegeois-Chauvel, et al., 2004; Purcell, et al., 2004) and pure tones (Maiste and Picton, 1989; Picton et al., 1987). Envelope frequencies between 4 and 16 Hz contribute the most to speech intelligibility (Drullman, et al., 1994a,b; van der Horst, et al., 1999). When provided with some simple spectral context (e.g., in frequency limited bands), the envelope alone is sufficient for speech to be intelligible. A set of four frequency limited bands of noise, modulated by the slow temporal speech envelopes in four matching frequency bands, can provide enough speech information to support near-perfect identification of vowels, consonants, and sentences (Shannon et al., 1995). Even without any spectral information—with envelope-modulated white noise—sentence recognition is significantly above chance levels (Shannon, et al., 1995; Van Tasell, et al., 1987). Given the importance of the envelope for speech intelligibility, an evoked response that follows the speech envelope could be particularly useful for validating the neural encoding of intelligible speech.

Magnetoencephalographic (MEG) responses after the slow temporal envelope of speech were recorded in a study that used varying amounts of time compression (Ahissar et al., 2001). An envelope-following response was detected by cross-correlating the scalp recorded response with the stimulus envelope, and this response was found to be correlated with comprehension. With increasing time compression, comprehension decreased, and the envelope-following responses were smaller. Electroencephalographic (EEG) responses to the speech envelope have been more difficult to recognize. Oygarden (2005) cross-correlated EEG signals with the envelopes of continuous speech and music stimuli but found only low correlations of borderline significance.

The present study recorded evoked potentials to the slow temporal envelope of spoken sentences. To improve the signal to noise ratio over ongoing EEG recordings, we averaged the responses over multiple presentations, and derived source waveforms from multichannel scalp recordings. Because the envelope frequency is too variable to be followed using the Fourier Analyzer approach that we used to evaluate responses to the glottal pitch, we used a cross-correlation procedure to detect responses. The cross-correlation approach to measuring responses has a long history in neurophysiology (e.g., Dobie & Wilson, 1984). The significance and latency of the maximum correlation indicate the presence and latency of the brain's response to the stimulus. Cortical responses to six different sentences were recorded and these responses were cross-correlated with the envelopes of the sentences.

There are multiple ways in which the brain might follow the speech envelope. A neural response synchronized to envelope modulations creates a temporal analogue of the envelope, which can be detected by cross-correlating the response with the stimulus envelope. The brain can also encode the envelope by increasing the firing rate of neurons tuned to modulation frequencies present in the envelope, but this type of encoding does not preserve the temporal structure of the envelope and therefore cannot be detected by cross-correlation or other temporal analyses of the scalp-recorded waveform. Both rate and temporal coding occur for the amplitude envelope at frequencies similar to those found in speech (Bieser & Mueller-Preuss, 1996; Liang, et al., 2002; Lu, et al., 2001; Schreiner & Urbas, 1986).

Because cortical neurons are very responsive to onsets, the brain could also follow the speech envelope by exhibiting a series of transient responses to successive envelope fluctuations (Draganova, et al., 2002; Phillips, et al., 2002). Both transient and ongoing responses exist in the auditory system. Neurons in the auditory cortex of the awake monkey tend to exhibit sustained or phase-locked (i.e., to the stimulus envelope) responses to stimuli for which they are optimally tuned, and onset responses for stimuli to which they are not tuned (Lu, et al., 2001; Wang, et al., 2005). Transient speech-evoked responses are characterized by overlapping N1-P2 or P1-N1-P2 complexes to onsets, offsets, and other acoustic changes, such as CV transitions (Agung, et al., 2006; Ostroff, et al., 1998; Tremblay, et al., 2003). The latter are called "acoustic change complexes." Because it would be difficult to determine what such a response would look like to the myriad acoustic changes in an entire sentence, we created a simple model of the expected response by convolving the grand average P1-N1-P2 complex to the onset of the sentence with the derivative of the stimulus envelope—the simplest measure of acoustic change. This model of the expected response was then cross-correlated with each of the source waveforms.

Materials and Methods
Subjects
Nine women (20 to 33 yr) and one man (24 yr) participated in the study. All participants were native English speakers, and had hearing thresholds of 20 dB HL or better from 250 to 8000 Hz (re American National Standards Institute S3.6 1996). All participants were right-handed.

Stimuli

Sixty sentences were created by appending 10 unique completions to six sentence roots, selected from the work of Bloom and Fischler (1980). The congruence of the completions was varied to engage participants in a task (to decide whether each sentence made sense), so that each sentence root could be repeated a number of times while still maintaining attention. Participant attention could also be verified by the presence of an N400—a negative deflection (when recorded with a noninverting electrode on the central-parietal region of the scalp) that occurs approximately 400 msec after the onset of a word that is not congruent with its immediate semantic context (Kutas & Hillyard, 1980). Selected sentence roots were 10 syllables in length, with at least five possible completions with cloze probabilities exceeding 0.03. The cloze probability associated with a particular sentence completion is the probability that the completion would be produced by a subject asked to finish the sentence, when provided only with the sentence root. A cloze probability of 0.03 indicates that 3% of respondents provided that completion. An additional set of five incongruent endings (completions had not been provided by any respondents) were prepared for each sentence, so that each sentence had five congruent and five incongruent endings. For example, the first sentence root was "To find the body they had to drain the" and sample congruent and incongruent endings were "lake" and "window." Sentence roots ranged in length from 1.64 to 2.64 sec, and each ending was approximately 1 sec long.

Sentences were spoken by a male speaker in a double-walled sound-attenuating chamber, and recorded using a Grason-Stadler 16 audiometer routed to a laptop computer. Sentences were digitized at 32 kHz with 16 bit resolution and recorded to a hard disk using Adobe Audition ™ (San Jose, CA). Each sentence root was combined with each of its 10 endings by splicing the final word into the sentence. Each combination was played aloud to ensure that the transition from the sentence root to the completion sounded natural, with no audible clicks. All sentences were scaled to have equal root-mean-square amplitude.

Speech Envelope

Each sentence root was filtered between 88 and 5657 Hz using a zero-phase implementation of a finite impulse response filter (>100 dB/octave). The envelope was calculated by computing the magnitude of the Hilbert transform (which provides the instantaneous amplitude of the signal). It was then low-passed filtered at 100 Hz and resampled at 250 Hz (the sample rate of the EEG response). The result was filtered between 2 (8 dB/octave) and 20 Hz (>100 dB/octave), using a zero-phase finite impulse response filter. Because psychophysical and electrophysiological responses generally vary in proportion to the log of stimulus magnitude, each envelope was transformed by taking 20 times the base 10 logarithm of the envelope. Figure 1 illustrates this process.
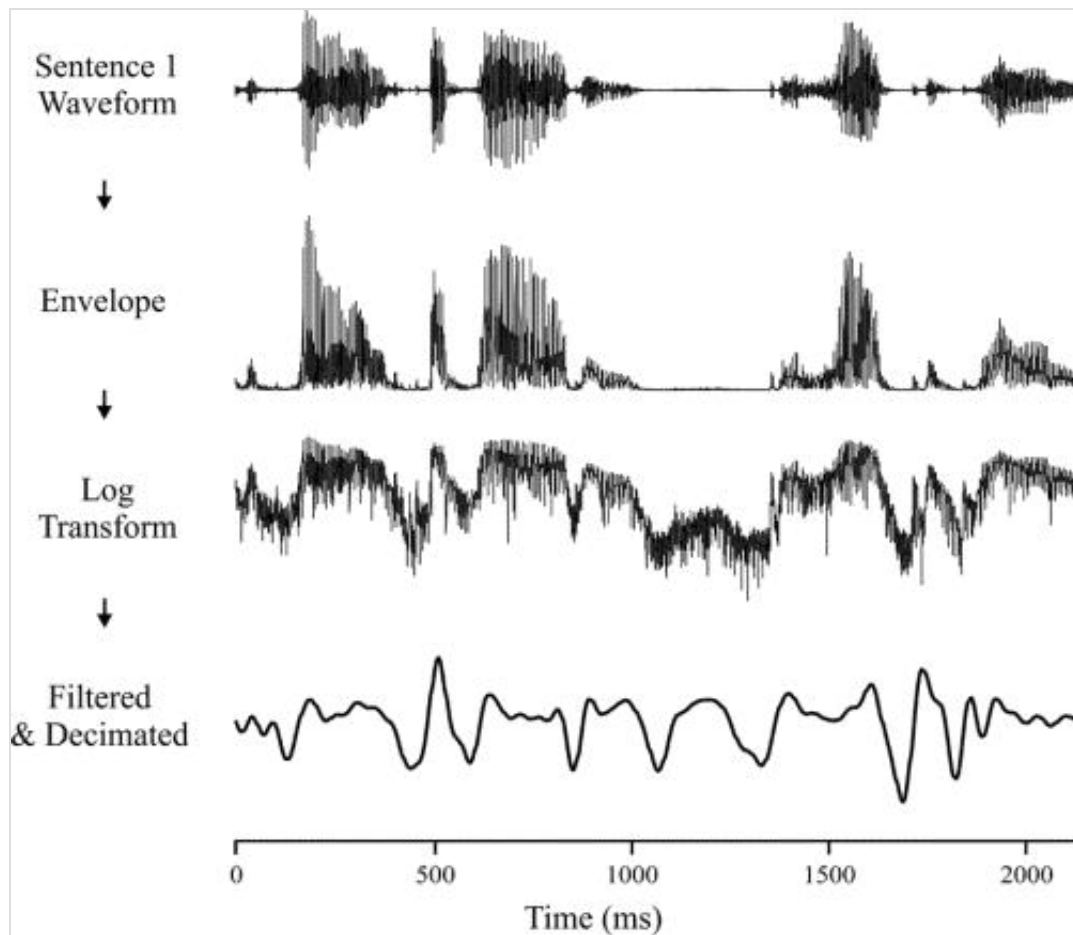
Fig. 1. Top, Waveform of the first sentence root ("To find the body, they had to drain"). Second, Temporal envelope of sentence. This was calculated by finding the absolute value of the complex frequency representation of the stimulus (produced using the Hilbert transform). Third, The logarithm of the envelope. Bottom, The log-envelope after filtering it between 2 and 20 Hz.

Procedure

A Tucker-Davis Technologies RP-2.1 system controlled by Matlab (Mathworks, Natick, MA) was used to present stimuli to the participant and send triggers to a Neuroscan SynAmps data acquisition system. The acoustic output of the Tucker-Davis Technologies system was routed through a GSI 16 audiometer, and presented binaurally via ER-3A insert phones. Sentences were presented so that their overall level was 60 dB SPL in a 2-cm^3 coupler.

Participants were seated comfortably in a double-walled sound-attenuating chamber and instructed to fix their gaze during stimulus presentation. One hundred twenty sentences were presented in each of five blocks, in a fully randomized order. This included 10 iterations of one congruent and one incongruent completion of each of the six sentence roots in each block, so that each sentence root was heard 100 times during the experiment. Participants were instructed to indicate whether each sentence "made sense" or "did not make sense" by pressing one of two response buttons. The interval between the offset of one sentence and the beginning of the next was fixed at 2 sec. Each block took approximately 12 min to complete.

Recordings

Participants were fitted with an electrode cap with 56 Ag/AgCl electrodes arranged in accordance with the International 10–20 system. The cap electrodes were FP1, FPz, FP2, AF7, AF3, AFz (ground), AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FC5, FC1, FCz, FC2, FC6, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP1, CPz, CP2, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, CB1, PO3, POz, PO4, CB2, O1, Oz, O2, and Iz. Off-cap electrodes were placed on the left and right mastoids (TP9/TP10), and on the sides of the face (FT9, FT10, F9, F10). Eye movements were recorded with electrodes placed on the infra-orbital ridge (IO1/IO2) and at the left and right outer canthi (LO1/LO2). Interelectrode impedances were kept below 5 kOhms. A Neuroscan Synamps system was used to acquire the electrophysiologic response at a rate of 250 samples per second. The response was amplified 500 times (least significant bit = 0.168 μV) and filtered between 0.15 and 50 Hz. All recordings were referenced to Cz during recording, and transformed to an average reference offline.

Separate recordings of vertical and horizontal eye movements, as well as blinks, were obtained before and after the experiment, and used to derive ocular source components (Ille, et al., 2002; Picton, et al., 2000) that were removed from each recording in Brain Electrical Source Analysis 5.16 (Berg & Scherg, 1994). In addition, trials with electrical activity exceeding 200 μV in any electrode (except the peri-ocular electrodes) were excluded from the averaging. This rejection process excluded 4.6% of the trials. Recordings for each sentence root were averaged separately for each participant.

Source Analysis

Source analysis was based on the first 3 sec of the grand average response to each sentence. This duration was short enough to exclude any cortical activity related to the button press (the shortest sentence root was 1.64 sec long before the final word, which itself lasted approximately 1 sec). Responses to each sentence were then concatenated to create a single grand average response to all of the sentence roots.

Source analysis was conducted in Brain Electrical Source Analysis 5.16, using the residual variance and energy criteria. The residual variance criterion seeks to minimize the residual variance in a model by iteratively adjusting the location and orientation of equivalent dipoles. The energy criterion seeks to minimize the total amount of activity attributed to the sources. Electrodes immediately adjacent to the eyes (IO1, IO2, LO1, LO2) were excluded from the source analysis because the ocular source components technique did not completely remove all the electrical activity caused by eye movements. Symmetrical regional sources were used to model the activity in each hemisphere, after filtering the responses between 2 and 20 Hz (12 dB/octave). Each regional source contains three orthogonal dipoles that can account for all the activity attributed to a particular location. Because most of the regional-source activity was attributed to two of the three dipoles, each regional source was converted into two single equivalent dipoles with identical origins. This removed the orthogonality constraint inherent in the regional sources, allowing the dipole orientations to be adjusted to model the response more precisely while maintaining the symmetry between the hemispheres. The largest dipole source in each hemisphere was oriented vertically, and the second dipole source was oriented horizontally. Figure 2 shows the location and orientation of the dipoles. Sources were fit to the concatenated grand average response that linked the responses (averaged over the 10 subjects) to each of the six sentences. After finding the best fit for the concatenated grand average response, second and third fittings were calculated for the initial segment of the response (0 to 300 msec) and the remaining portion (300 to 3000 msec). These fits were not significantly different from the initial solution. The source dipoles calculated on the concatenated grand average response solution (Fig. 2) were therefore adopted for the study. Source waveforms were computed for each source dipole, each sentence, and each subject, and exported to Matlab. Source waveforms show the changes over time in the current at each dipole source. For each subject and each sentence we therefore obtained four source waveforms, two (vertical and horizontal) in each hemisphere. Each source waveform was then filtered between 2 (8 dB/octave) and 20 Hz (>100 dB/octave), using a zero-phase finite impulse response filter. Figure 2 shows the source waveforms for a single sentence.
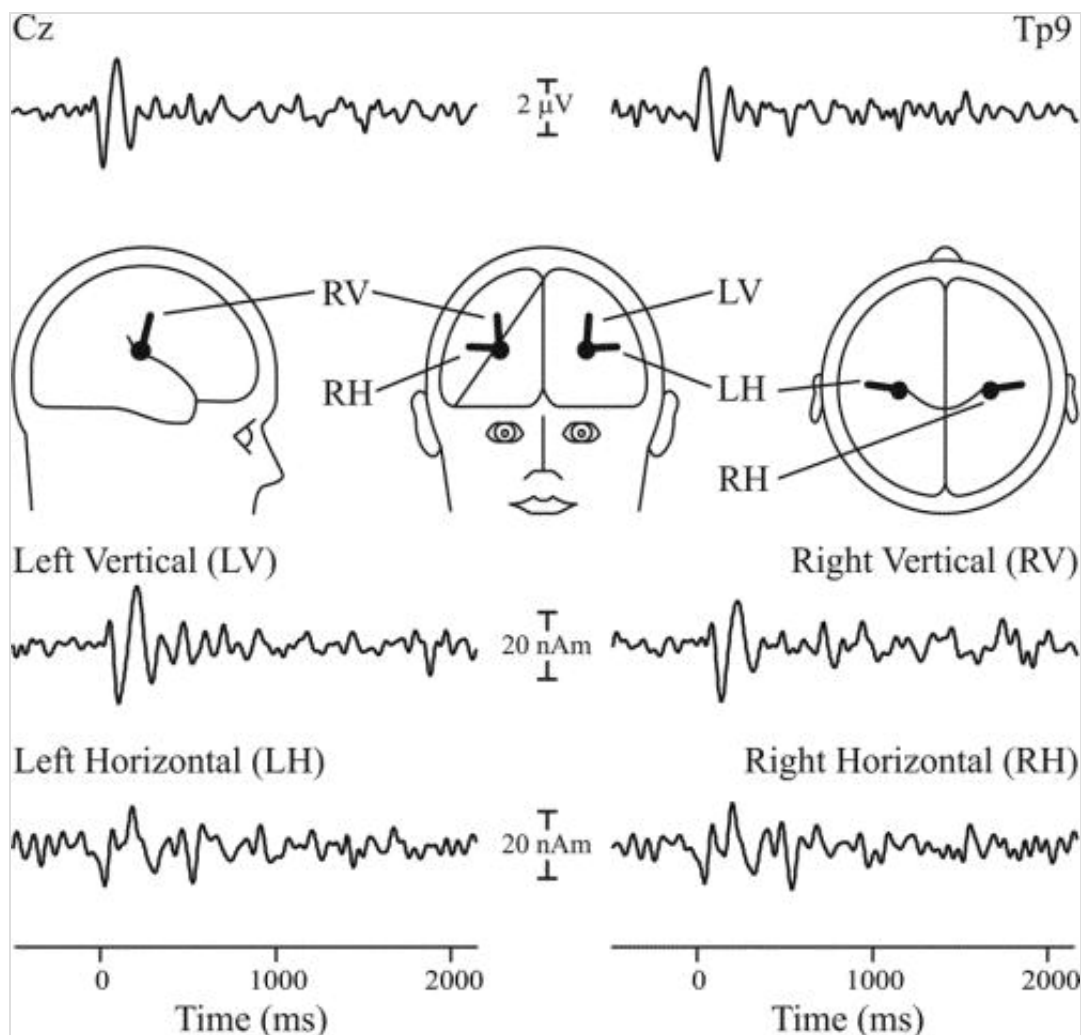
Fig. 2. Top, Responses to the first sentence at the Cz and TP9 electrodes. The activity at the left mastoid (TP9) is largely the inverse of activity measured at Cz, reflecting the orientation and position of the dipole in the auditory cortex. Second, Diagram showing locations of horizontal and vertical dipoles in the sagittal, coronal, and transverse planes. Third, Source waveforms for the first sentence at the left and right vertical dipoles. Fourth, Source waveforms for the first sentence at the left and right horizontal dipoles. Positive plotted upward.

Cross-Correlations

To determine whether responses were related to the stimulus envelope, source waveforms and sentence-root envelopes were compared using windowed cross-correlation. The windowed procedure allowed for the possibility that the correlations might change across different time periods within the sentence. A window of 500 msec was chosen as the shortest window that could include a full cycle of the lowest envelope frequency (2 Hz). The correlation between the stimulus envelope and the source waveform was computed in successive 500 msec windows beginning 250 msec before the onset of the sentence and continuing at 4 msec intervals through the sentence. Each 500 msec window of the stimulus waveform was correlated with 500 msec windows of the source waveform (response) occurring at delays ranging from 0 to 300 msec (again in 4 msec intervals). The delay at which the correlation was greatest (or least) could be assessed at each time in the stimulus envelope.

The correlations were plotted using the "jet" color scale (see Figure 3), where warm colors (red and yellow) indicated a positive correlation and cool colors (light and dark blue) indicated a negative correlation. Correlations were plotted with respect to the center of each window (in stimulus time) on the x axis, and with respect to the delay between the stimulus envelope and the response on the y axis. A consistent positive correlation between the stimulus envelope and the response at a particular delay would be represented by a horizontal band of warm color, with the vertical position of the band indicating the delay of the response.
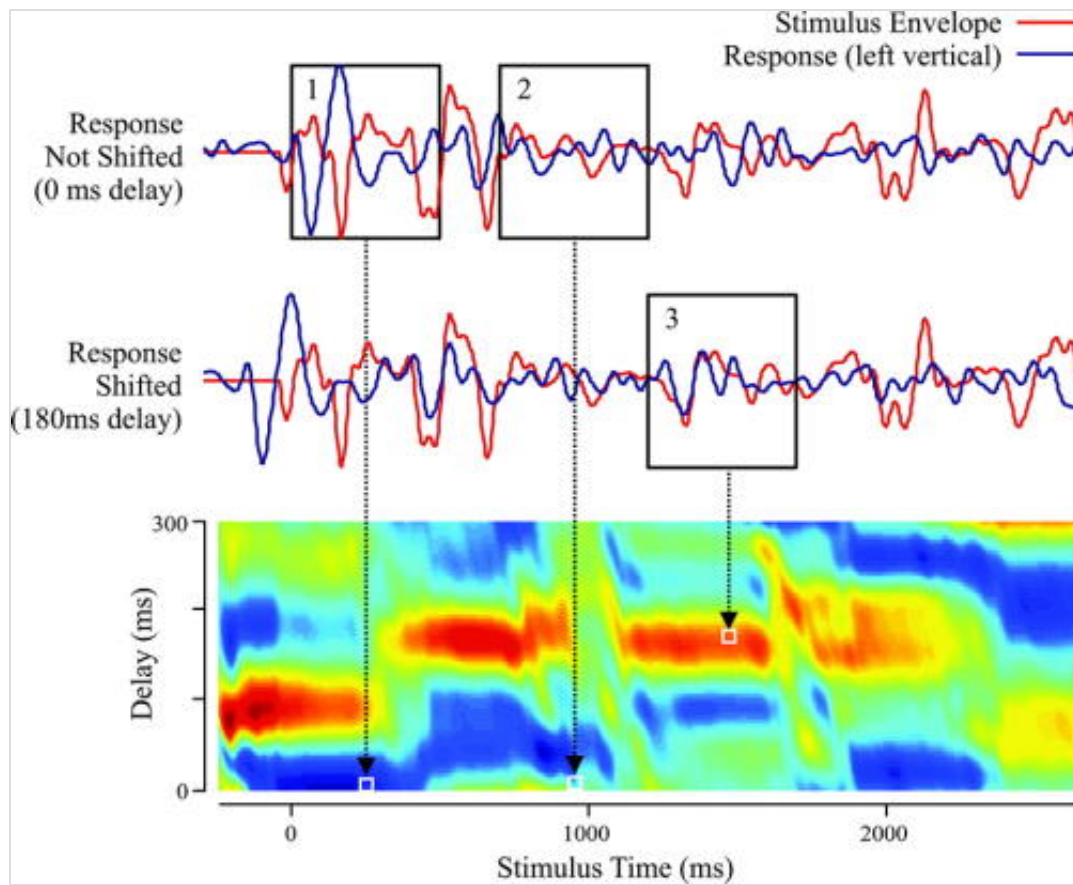

Fig. 3. Top, The filtered-log envelope of sentence 1 (red) and the average source response at the left vertical dipole (blue). The correlation between these two signals was calculated in 500 msec-wide windows at every sample point (i.e., 250 times per second). These correlations are plotted in the bottom row of the correlelogram (bottom), using the "jet" color scale. Two windows are marked. In the first window, the sentence envelope and the response seem to be negatively correlated. This is reflected in the correlelogram with a blue dot. In the second window, the stimulus and response are neither negatively nor positively correlated. This is reflected with a green dot in the correlelogram. Second, The same log-envelope of the stimulus and the response, after shifting the response forward 180 msec (to account for the possibility that the response may have occurred 180 msec after the stimulus). A third window is marked. In this window, the envelope and the response are positively correlated. This is reflected in the correlelogram with a red dot. Bottom, The correlelogram showing the correlations between the envelope and response at each point in stimulus time (x axis), and at response delays from 0 to 300 msec (y axis). The color scale is automatically adjusted to span from the minimum correlation ($r = -0.76$) to the maximum correlation ($r = 0.84$).

Although the magnitude of the correlation might vary across time, any true correlation between the stimulus and response (or the transient model and the response) would need to be consistent over time. Spurious correlations between stimulus and response would not likely occur at a consistent delay. One way to evaluate the consistency of the correlations is to take the mean correlation over stimulus-time at each response delay. The expected value of exclusively spurious correlations is zero. Figure 4 (upper right) shows the result of averaging the correlations at each response delay (i.e., in each row of the upper left plot). The 250 msec of the response immediately after the beginning of the sentence was excluded from this average, because this period was characterized by a large onset response. Note that after averaging, there is one large negative correlation at a response delay of 120 msec, and a larger positive correlation at 188 msec. This would suggest that the response followed the stimulus envelope at 188 msec, and followed the inverse of the envelope at 120 msec. The average correlation has been plotted in the second row of the figure with delay on the x axis. The fast Fourier transform (FFT) of the average correlation gives the frequencies that are contributing to the correlations. This was performed after zero-padding the 300 msec waveform out to 1 sec to improve the frequency resolution.
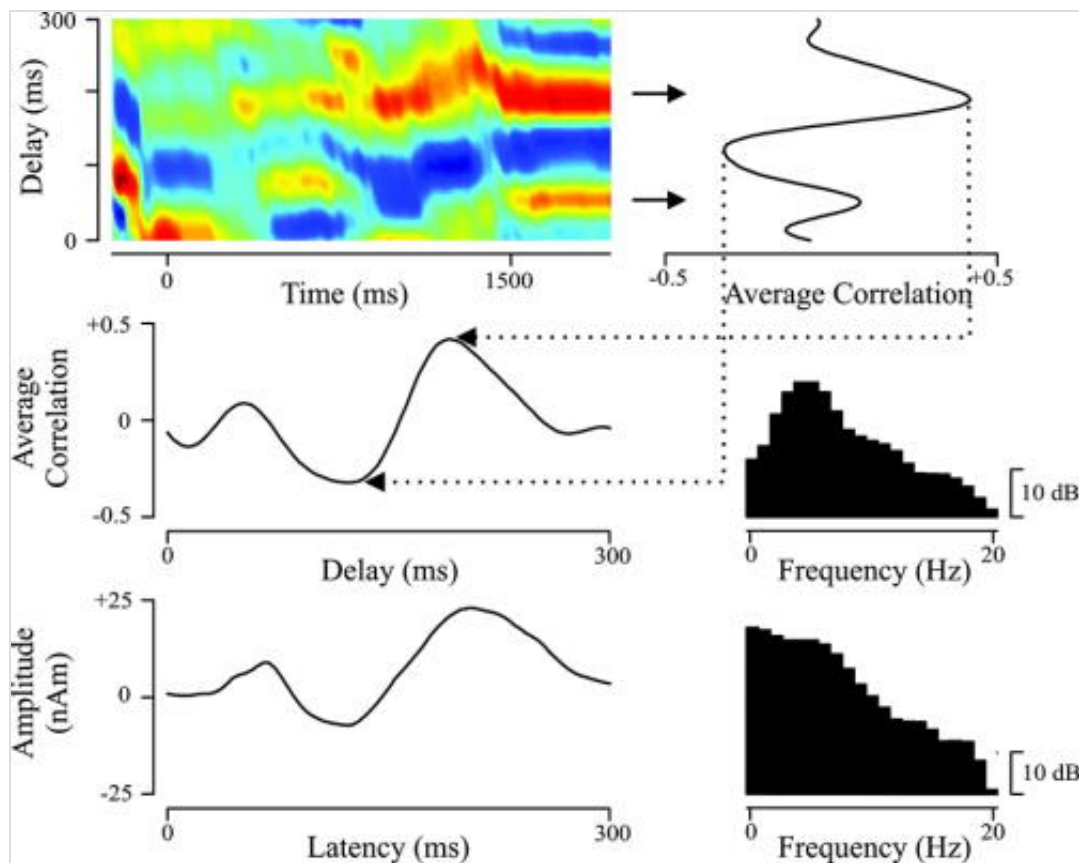

Fig. 4. Top left, Correlelogram showing correlation strength with respect to stimulus time (x axis) and response delay (y axis), for sentence 1. The large positive correlation (at a delay of approximately 200 msec) begins approximately 300 msec after the onset of the sentence, likely because the period immediately following the onset is dominated by a large transient response (bottom left). Top right, The average correlation across time, plotted with respect to delay (y axis) and correlation strength (x axis). The average excludes the 250 msec immediately following the sentence onset. Middle left, The average correlation inverted so that delay is on the x axis, and correlation strength is on the y axis. This orientation is adopted for use in later figures. Middle right, Average spectrum of correlations between all sentence envelopes and all sources in the auditory cortex, obtained after zero-padding the average correlation waveform out to 1 sec. The dominant frequencies in the correlations fall between 2 and 7 Hz (the 6-dB down points), with a peak at 4 Hz. Bottom left, Response to the sentence onset, averaged across all sentences and all participants, at the two vertical dipoles. Note that this is plotted with respect to latency from the sentence onset. Bottom right, Spectrum of transient onset response (obtained after zero-padding out to 1 sec). If the onset response is treated as the impulse response of a cortical "filter," the spectrum is equivalent to the transfer function of the filter.

The significance of the mean correlation can be assessed by determining the likelihood that the mean correlation could have occurred by chance. If the mean correlation exceeds more than 95% of the correlations that would be expected to occur by chance alone, it can be assumed to be significant at $p <$ 0.05. The distribution of correlations that could occur by chance was determined by randomly reordering the correlations with respect to delay (because accidental correlations should not occur at a consistent delay), at each point in stimulus time, before calculating the mean correlation. This process was repeated 2400 times (100 times for each sentence and source), to generate a stable distribution. Because correlations can be negative or positive, the mean of random correlations (across all stimulus time windows) should be smaller than the mean of correlations occurring at a consistent delay. It is the same principle that motivates the averaging of a large number of evoked responses to detect small potentials that would otherwise be buried in the background electrical activity.

The validity of this approach rests on the assumption that random correlations do not occur at a consistent delay. Because the brain produces rhythmic activity even in the absence of stimulation, a random correlation between the brain response and the speech envelope might sometimes occur at a consistent delay. Rhythmic activity in the Delta band (2 to 3 Hz), Theta band (4 to 7 Hz), Alpha band (8 to 12 Hz), and Beta band (13 to 20 Hz) could correlate with the speech envelope between 2 and 20 Hz, even if no true relationship existed. Randomly reordering correlation values with respect to delay could eliminate an ongoing accidental correlation between a cortical rhythm and the stimulus envelope, and thereby underestimate the magnitude of the average correlation that could occur by chance. To assess this possibility, we also calculated the average correlation between each stimulus envelope and scalp-recorded activity during periods where there was no stimulation. Because the interstimulus interval was 2 sec long, and any response to the stimulus should have occurred within the first second, the last second of each interstimulus interval was used for this purpose. An alpha of 0.05 was assumed for all statistical analyses.

Transient Response Model

An alternative model of the envelope-following response (hereafter referred to as the "transient response model") was also created by convolving a typical P1-N1-P2 response with the stimulus envelope derivative. This model assumes that a P1-N1-P2 complex is continuously elicited by changes in the stimulus envelope. The P1-N1-P2 complex used in the model was the average response to the beginning of each sentence (from 0 to 300 msec), from the left and right vertical source waveforms. It is displayed in the bottom left panel of Figure 4. Peak amplitudes and latencies are indicated in Table 1. The latencies of the peaks were 68 msec (P1), 124 msec (N1), and 208 msec (P2).

TABLE 1. Amplitudes and latencies of grand average onset response to all sentences (vertical sources)

	P1	N1	P1
Amplitude (nAm)	9.1	7.3	23.2
Latency (msec)	68	124	208

TABLE 1. Amplitudes and latencies of grand average onset response to all sentences (vertical sources)

The convolution process multiplied each point in the stimulus envelope by each point in the transient response and summed the resultant products:

$$y(t) = \sum_{k=0}^{N} x(k)h(t - k)$$

Equation (Uncited)

where y(t) is the convolved waveform, h(k) is the transient response and x(t) is the stimulus envelope. Because convolution with a negative derivative value would result in an inversion of the P1-N1-P2, the derivative was made exclusively positive by adding the absolute value of its maximum negative deflection. The convolved waveform of length N (equal to the sum of the lengths of the two waveforms being convolved minus 1) was truncated to the length of the envelope. The convolved waveform (i.e., the transient response model) was then compared with all response source waveforms using windowed cross-correlation. The results of these procedures are shown in Figure 5. The convolution process is basically the same as a filter. In this context, the transient P1-N1-P2 complex can be considered the impulse response of the filter. As such it can be analyzed to give a transfer function, from which we can derive the dominant frequencies that it passes, and the group delay over these frequencies. The magnitude of the transfer function is displayed in the bottom right panel of Figure 4.
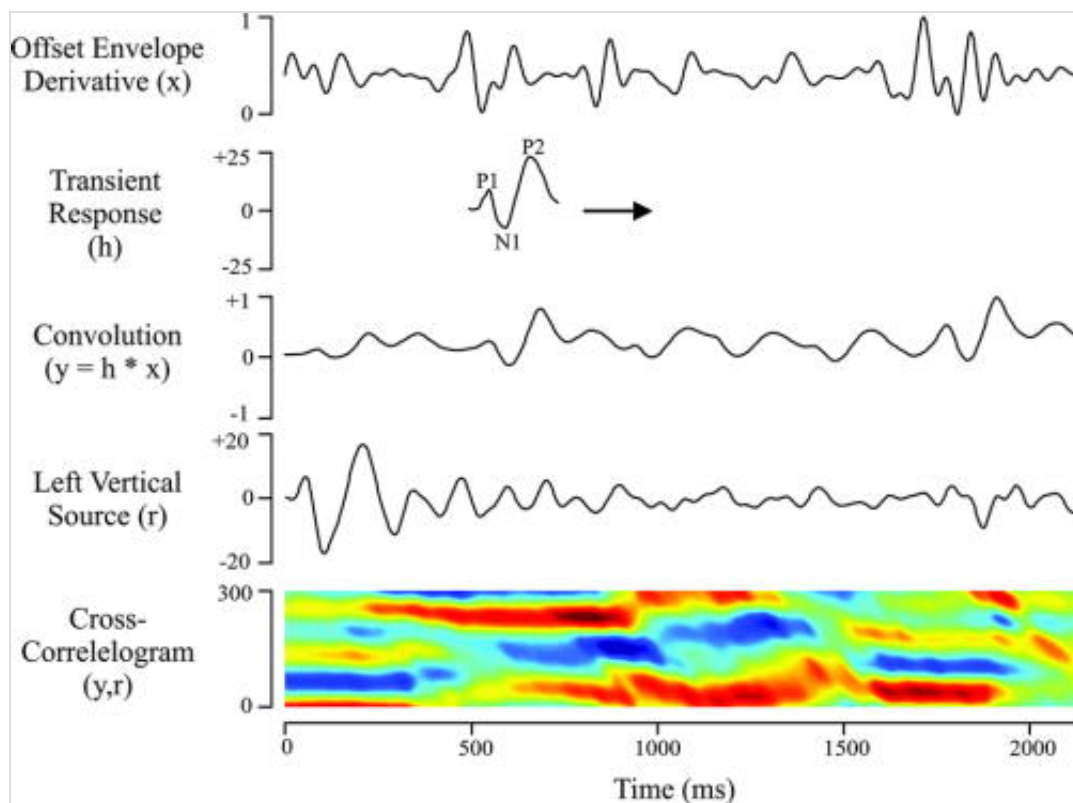

Fig. 5. Transient response model. Top, Derivative of the log-envelope of the first sentence, offset to be above zero. Range has been normalized by dividing by maximum value. Normalization has no effect on the cross-correlation. Second, The average transient response to the onset of all sentences. This is characterized by three peaks—a P1 at 68 msec, an N1 at 124 msec, and a P2 at 208 msec. Values are in nAm. Third, The convolution of the log-envelope derivative (shifted to be above zero) of the first sentence and the transient onset response, truncated to the length of the derivative, and normalized by dividing by maximum absolute value. The transient response is moved sequentially through the envelope derivative to form the convolution. The convolution is the sum of a series of transient responses multiplied with each successive element of the positive envelope derivative. Thus peaks in the derivative tend to give a convolution waveform that is similar to the transient response. This is the transient response model for the first sentence. Fourth, The average response (left vertical source) to the first sentence, in nAm. Fifth, Correlelogram showing the correlation between the transient response model and the response. Values are plotted with respect to stimulus time in msec (x axis) and delay in msec (y axis), using a color scale adjusted to span from the minimum correlation (r = -0.80) to the maximum correlation (r = 0.75). There is a band of increased correlation though most of the sentence at a delay of 0 to 50 msec.

Results
Scalp and Source Waveforms

The scalp-recorded responses to the sentences showed a large P1-N1-P2 complex at the beginning of the sentence followed by a sequence of smaller fluctuations. These responses were largest over the fronto-central regions of the scalp and inverted in polarity at the mastoid and other posterior-inferior electrodes (see top row, Fig. 2). On FFT analysis, the average responses showed their greatest energy below 10 Hz. Figure 6 presents the average spectrum of the response (across all sentences and subjects) at the Cz and Oz electrodes (third row), and compares this with the average spectrum of stimulus envelope modulations (top row). In addition, the ongoing spectra of the EEG at Cz and Oz are shown (second row). These spectra are averaged single-trial spectra. Alpha activity (8 to 12 Hz) is more prominent in the spectrum of the EEG, particularly at the Oz electrode. This has been cancelled out in the averaging process, so the spectrum of the average response is similar to the spectrum of the speech envelope.
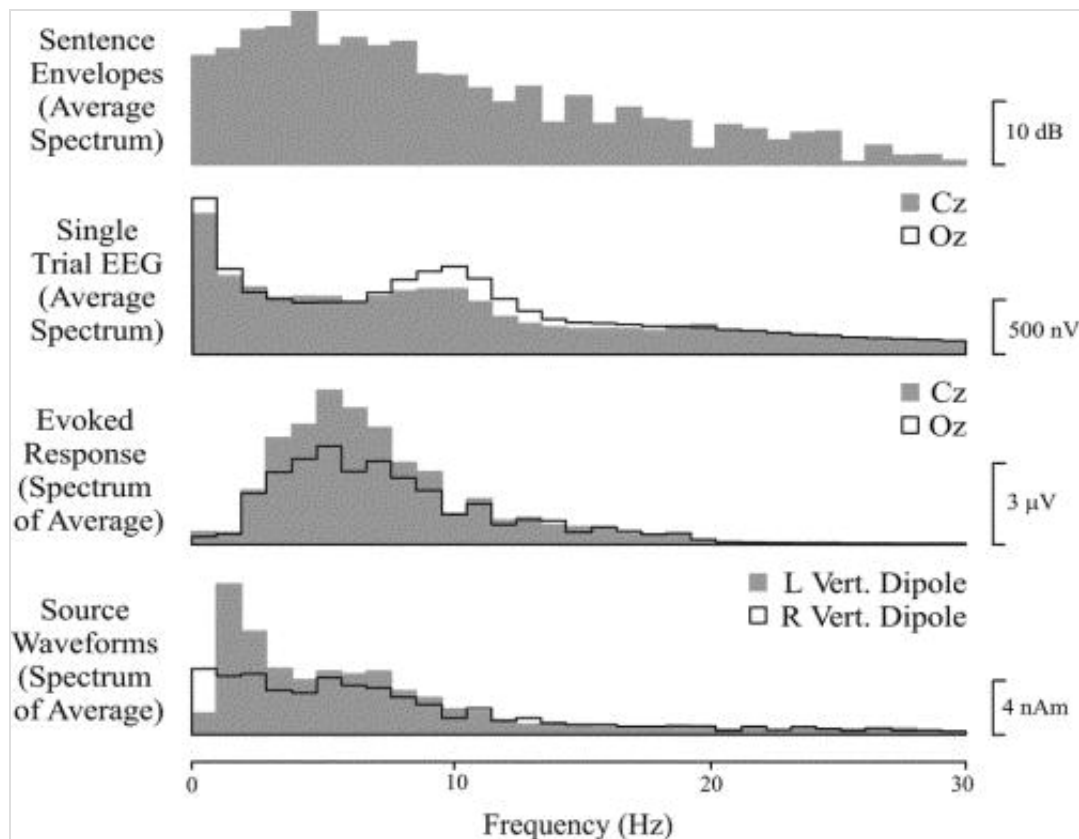

Fig. 6. Top, Average of all sentence envelope spectra. Second, Average of all individual trial spectra at Cz and Oz electrode locations. This shows periodic activity that is both time-locked and nontime-locked to the stimulus. The peak near 10 Hz reflects the alpha rhythm. Third, Average spectra of the averaged evoked response to each sentence, at the Cz and Oz electrode locations. Spectra calculated after averaging show only time-locked activity, so the alpha rhythm cannot be seen. Bottom, Average spectra of the averaged response, projected onto the left and right vertical source dipoles.

The sources for the scalp waveforms were localized to the posterior region of the Sylvian fissure, in the vicinity of the planum temporale. The sources were modeled as a pair of vertical and horizontal dipoles, symmetrical in location and mirror-image in orientation. Figure 2 shows the location and orientation of the dipoles, as well as the grand average source waveforms to the first sentence. The spectrum of the activity at the left and right vertical sources is displayed in the bottom row of Figure 6.

When responses to sentences with congruent endings were subtracted from responses to sentences with incongruent endings, we observed an N400—a small negative deflection beginning 400 msec after the onset of the final word, maximal over the central-parietal electrodes. This indicates that participants were paying attention to the sentences.

Cross-Correlations with Sentence Envelope

The grand average cross-correlations between each sentence and the left and right vertical and horizontal dipole sources are plotted in Figure 7. Cross-correlation peaks between the vertical and horizontal source waveforms, and all sentences ranged from 152 to 208 msec. For the left vertical source, the average short-term cross-correlation was highest ($r = 0.37$) at a delay of 175 msec, and for the right vertical source, the average correlation was highest ($r = 0.33$) at a delay of 180 msec. For the left horizontal source, the average cross-correlation was highest ($r = 0.22$) at a delay of 165 msec, and for the right horizontal source, the average correlation was highest ($r = 0.28$) at 163 msec. Average correlations with the right and left vertical sources are displayed in Figure 8 (top). We also calculated cross-correlations for each subject averaged across all sentences. These are displayed in Figure 8 (bottom). For individual subjects the average correlation between the envelope and the left vertical dipole was highest ($r = 0.16$) at a delay of 177 msec, and the average correlation between the envelope and the right vertical dipole was highest ($r = 0.18$) at a delay of 185 msec. The average correlation with the left horizontal dipole was highest ($r = 0.14$) at a delay of 165 msec and with the right horizontal dipole was highest ($r = 0.15$) at a delay of 164 msec. Peak amplitudes and latencies of responses averaged across sentences and averaged across participants are displayed in Table 2.
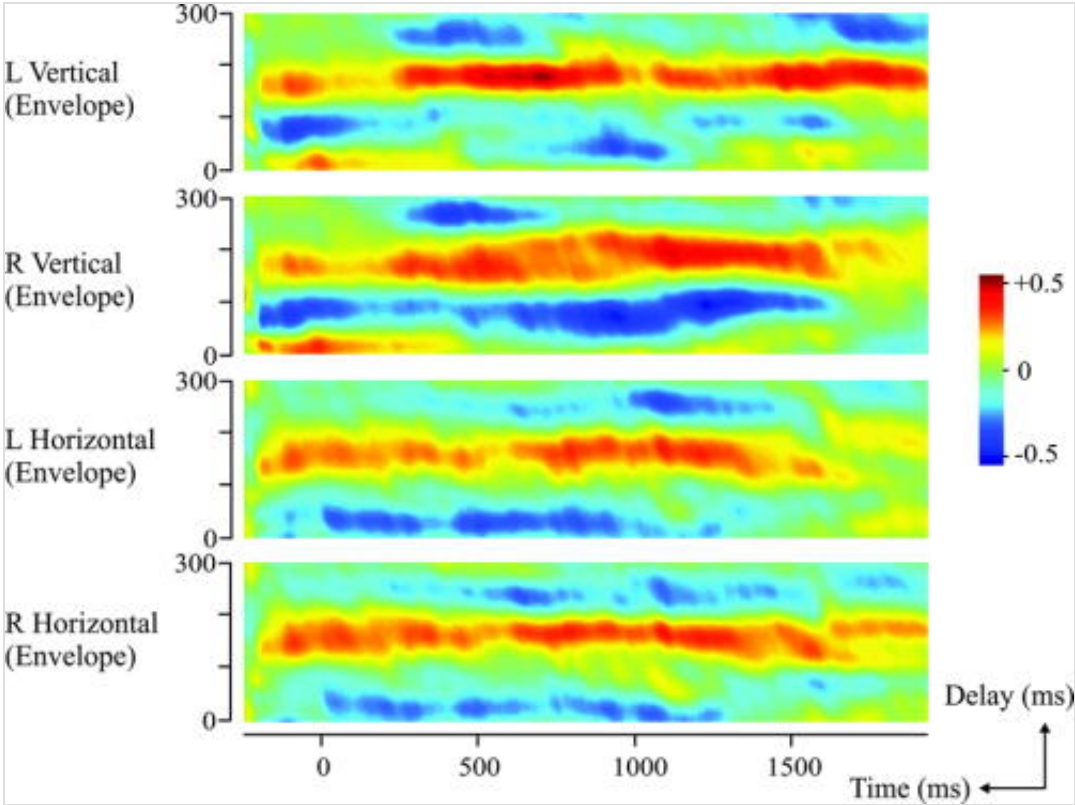

Fig. 7. Average correlelograms showing the magnitude of correlation between sentence envelopes and all source waveforms, plotted with respect to stimulus time (x axis) and response delay (y axis). Correlations are plotted using the "jet" color scale, where warm colors reflect positive correlation, and cool colors reflect negative correlation. Bands of warm color in the center of each chart indicate that all the sources followed sentence envelopes at delays ranging from 150 to 200 msec.
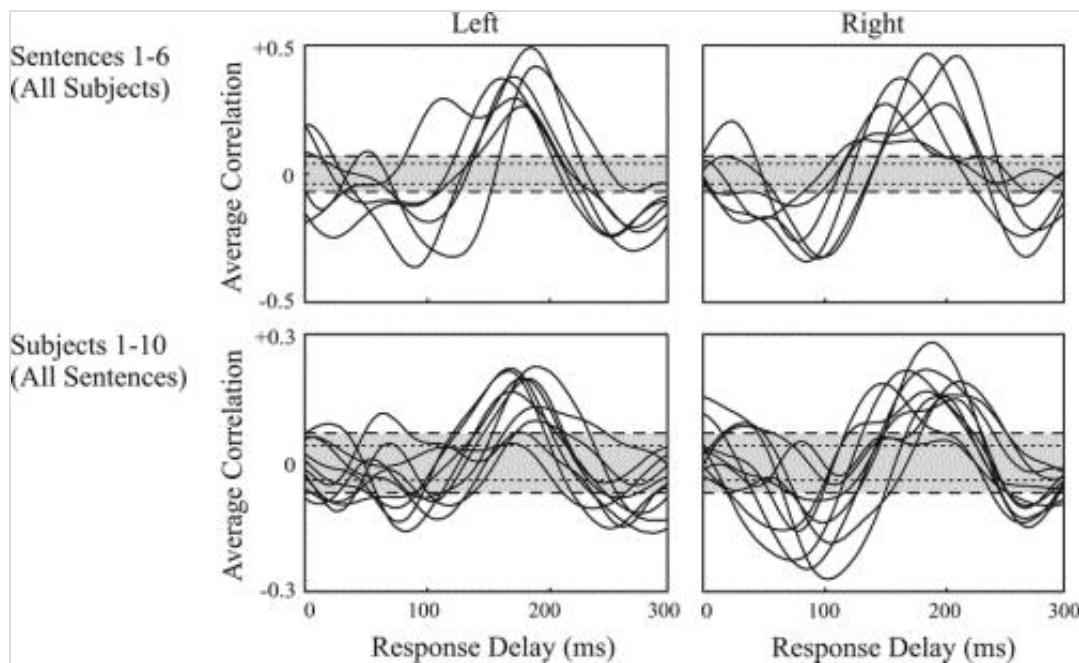
Fig. 8. Average correlations between log-envelopes and responses at left and right vertical dipoles for each sentence (top) and participant (bottom). The largest correlations occurred at a delay of approximately 180 msec. The dotted lines (within the gray area) on each chart reflect the range which encompasses 95% of all correlation values that would be expected to occur by chance, as determined by cross-correlating envelopes with responses that have been scrambled (in 2400 permutations). The dashed lines that border the gray area reflect the range which encompasses 95% of all correlation values that occurred in the absence of stimulation. This range was determined by correlating sentence envelopes with responses recorded when no stimulus was being presented.

TABLE 2. Mean peak correlations and peak correlation latencies averaged across sentences and subjects, at vertical and horizontal sources

	Envelope – Response				Transient model – Response			
	L vertical	R vertical	L horizontal	R horizontal	L vertical	R vertical	L horizontal	R horizontal
Average Sentence								
Peak correlation (r)	0.37	0.33	0.22	0.28	0.31	0.30	0.20	0.21
Latency of peak (msec)	175	180	165	163	10	6	7	5
Number significant	6/6	6/6	5/6	6/6	6/6	6/6	5/6	6/6
Average Subject								
Peak correlation (r)	0.16	0.18	0.14	0.15	0.14	0.16	0.11	0.12
Latency of peak (msec)	177	185	165	164	14	14	3	2
Number significant	9/10	9/10	9/10	10/10	9/10	10/10	7/10	6/10

TABLE 2. Mean peak correlations and peak correlation latencies averaged across sentences and subjects, at vertical and horizontal sources

The significance of these correlations was assessed by comparing them to the distribution of correlations that would be expected to occur by chance, after randomizing with respect to delay. A correlation could be considered to be significant at $p < 0.05$ if its absolute value was greater than 0.04. This range is indicated in Figure 8 by the dotted lines within the gray shaded area. However, as noted above, the brain produces rhythmic activity in the absence of stimulation, which could result in accidental correlations with the speech envelope. Ninety-five percent of the correlations between the sentence envelopes and the portions of the response where there was no stimulus occurred when r was between 0.07 and -0.07. This range is indicated in Figure 8 by the shaded area bounded by dashed lines. Because this range is greater than the range calculated after randomizing delay, it seems that there was some correlation between the stimulus envelopes and cortical oscillations in the absence of stimulation. Using this response criterion, all grand average vertical sources, and all but one of the grand average horizontal sources (i.e., the left horizontal source to sentence 6) were significantly correlated with the sentence envelopes. Vertical sources were significantly correlated with sentence envelopes for all but one of the individual subjects. Right horizontal sources were significantly correlated with sentence envelopes for all subjects, and left horizontal sources were significantly correlated for all but one of the subjects.

There were also significant positive correlations between the right and left vertical sources and the stimulus envelope at delays between 0 and 72 msec, and significant negative correlations at delays ranging from 76 to 148 msec and 228 to 296 msec. The time differences between these positive and negative correlation peaks were related to the dominant energies in the speech envelope and the response (see Fig. 6). For example, in sentence 1, the first positive correlation maximum occurred at 52 msec. This was followed by a negative correlation maximum at 120 msec (68 msec later), and a positive correlation maximum at 188 msec (68 msec later). The length of time between these positive and negative correlations (68 msec) is equal to the half-wavelength of a 7.4 Hz frequency component—a dominant frequency in the envelope and the response. If the signal and the response were sinusoids, each negative and positive correlation would be expected to have the same absolute magnitude, because a half-wavelength shift is equivalent to an inversion of the sinusoid. Although speech envelopes are not sinusoidal, a delay shift equal to the half-wavelength of a dominant frequency would still likely produce a correlation, albeit with less absolute magnitude. The FFT of the correlations between all sentences and all sources (middle right of Fig. 4) showed the main frequencies in the correlation to be between 2 and 7 Hz (the 6 dB down points), with a peak at 4 Hz. Therefore, the smaller negative and positive correlation peaks (above and below the largest positive correlation) most likely reflected the dominant frequencies in the correlation rather than additional envelope-following responses occurring at different delays.

Cross-Correlations with Transient Response Model

The transient response model was determined by convolving the positive-only derivative of the stimulus envelope with the average P1-N1-P2 response to the onset of the sentences. This transient response was therefore like the impulse response of a filter. The analysis of this impulse response gave an amplitude transfer function that was basically low-pass with a cutoff at around 7 Hz. (bottom right of Fig. 4). The average group delay over the frequencies below 7 Hz was 209 msec.

We calculated the cross-correlations between the grand average source waveforms and the transient response models. Because these correlations were being calculated with models of the cortical responses, the peaks of the cross-correlations were expected to occur at short time delays. The correlelogram for the left vertical source is illustrated in the bottom of Figure 5. Average correlations for each sentence and each subject are displayed in Figure 9. Peak amplitudes and latencies are indicated in Table 2. For the left vertical source, the highest average correlation ($r = 0.31$) occurred at a delay of 10 msec and for the right vertical source, the highest average correlation ($r = 0.3$) occurred at a delay of 6 msec. For the left horizontal source, the highest average correlation ($r = 0.2$) occurred at a 7 msec delay and for the right horizontal source, the highest average correlation ($r = 0.21$) occurred at a 5 msec delay. The transient response model correlated significantly with right and left vertical sources for all of the sentences, and with right and left horizontal sources for all but one of the sentences (the left horizontal source for sentence 1). For individual subjects, the highest correlation with the left vertical source ($r = 0.14$) and the right vertical source ($r = 0.16$) occurred at a delay of 14 msec. The highest correlation with the left horizontal source ($r = 0.11$) occurred at 3 msec and the highest correlation with the right horizontal source ($r = 0.12$) occurred at 2 msec. The model correlated significantly with all the right vertical sources, and with all but one left vertical source. The model correlated significantly with both left and right horizontal sources for four subjects, with the left horizontal source for three subjects, with the right horizontal source for two subjects, and with no horizontal sources for one subject.
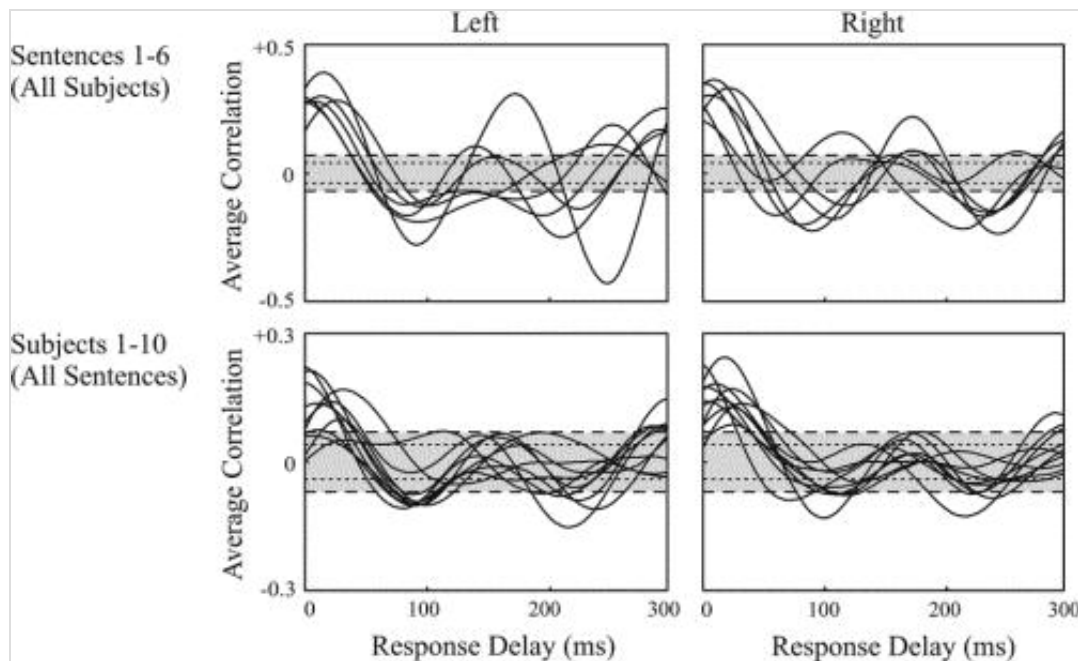

Fig. 9. Average correlations between transient response models and responses at the left and right vertical dipoles for each sentence (top) and participant (bottom). Dotted and dashed lines are same as in Figure 8.

Envelope and Transient Response Model Comparisons

Because the source waveforms were correlated with both the stimulus envelopes and with the transient response models, peak correlation magnitudes were compared using a within-subjects analysis of variance, with correlation type (envelope versus transient response model) and dipole as the independent variables. There were no significant interactions, but there was a main effect of the type of correlation ($F(1,9) = 33.6$, $p < 0.05$), such that the mean correlation with the stimulus envelope was significantly higher than the mean correlation with the transient response model.

Discussion
Evoked Potentials to Sentences

Each sentence evoked a large onset response and then a series of smaller deflections. Other studies of evoked potentials to sentences have not concerned themselves with these smaller deflections, usually averaging together responses to different sentences or looking at the responses to specific events within a sentence (Connolly, et al., 1990, 1992; D'Arcy, et al., 2004). Most event-related potentials studies of sentence processing have used visually presented stimuli which allow better time-locking of stimulus to response (e.g., Kutas & Hillyard, 1980; Yamada & Neville, 2007). Hemodynamic studies of sentence processing can localize activations but do not have the temporal resolution to show specific envelope effects (Zekveld, et al., 2006). Our paradigm averaged multiple responses to the same sentences while maintaining the subjects' attention by having them judge the congruence of the final word. This allowed us to use auditorily presented sentences and to focus on the small deflections that occurred during the sentence. The presence of an N400 after incongruent words indicates that the participants did indeed pay attention to the sentences.

Correlations with the Speech Envelope

The complex pattern of slow envelope fluctuations across the spectrum is essential to understanding speech. Because the human auditory cortex produces a synchronous response to regular modulations that are similar in frequency to the envelope fluctuations in speech (e.g., Maiste & Picton, 1989), it is reasonable to expect that cortex produces a synchronous response to the more variable modulations of the speech envelope. We found that vertical and horizontal source waveforms in the auditory cortex provide a representation of the stimulus envelope (for all sentences) shifted in time by a delay of 152 to 208 msec.

Correlations with the Transient Response Model

Although neurons in the cortex could lock onto and follow the speech envelope, and generate the response that we recorded, the response could also be related to the acoustic change complex that is typically recorded to acoustic changes during speech. Evoked potential studies of words or syllables often show both an onset response and subsequent smaller responses to the phonemic changes (e.g., Ostroff, et al., 1998; Tremblay, et al., 2006a). These later waves have been termed acoustic change complexes (Martin & Boothroyd, 1999). It is possible that the small deflections in the event-related potential that followed the onset response were not envelope-following responses, but rather multiple acoustic change complexes evoked by phonemic changes occurring during the sentence. Both the onset response and the subsequent waves showed a similar scalp topography and gave similar sources on source analysis. These scalp responses were largely generated in the superior regions of both temporal lobes with the equivalent dipoles localizing to the planum temporale. For this reason, we modeled the envelope response as a sequence of transient responses evoked as the envelope changed.

Source waveforms were well predicted by the convolution-based model. The model correlated best with vertical and horizontal source waveforms at delays from 0 to 24 msec, suggesting that the onset response to the sentence provided a reasonable estimate of the transient response pattern with respect to the log-envelope derivative. The latency of the P2 in the model (208 msec) was based on the grand average onset response to all the sentences, and was similar to the latency of the typical acoustic change complex P2 (200 to 220 msec), so the response could have been a series of acoustic change complexes.

But how might this be related to the envelope-following response? Although the envelope-following response occurred 152 to 208 msec after the envelope, the peaks of the derivative generally precede the peaks of the envelope. The temporal relationship between each sentence envelope and its derivative was determined by finding the peak of the cross-correlation function between each envelope and its derivative. In all cases, a sentence's envelope derivative preceded its envelope by 24 to 32 msec. Thus, the expected latency of the P2 with respect to the envelope derivative would be expected to be greater than the delay of the envelope-following response. Specifically, the latency of the P2 would be expected to fall between a minimum of 176 msec (152 msec + 24 msec) and a maximum of 240 msec (208 msec + 32 msec). The mean of this range is 208 msec, which is the same as the modeled P2 latency. The envelope-following response (delay: 152 to 208 msec) could thus have been equivalent to the P2 of the acoustic change complex (latency: 200 to 220 msec) and also to the P2 response convolved with the envelope derivative (latency: 208 msec). All three are roughly coincident when the temporal relationship between the envelope and its derivative is taken into account.

The Nature of the Cortical Response to the Speech Envelope

Neurons in the auditory cortex generally produce transient responses (increases in firing rate), as a function of the rate of change (the derivative) or acceleration (the second derivative) of the stimulus envelope (Heil, 1997a,b; Phillips, et al., 2002). There is a tendency for increased interneural synchrony while a steady state stimulus is being presented (Eggermont, 1997), but sustained neural responses are found only in the small subset of neurons that are optimally tuned to the stimulus (Lu, et al., 2001; Wang, et al., 2005). Because the scalp-recorded response reflects the summed postsynaptic activity across a large number of neurons, it could reflect small increases in neural synchrony, or sustained responses produced by a subset of optimally-tuned neurons, but it is most likely dominated by widespread transient activity synchronized to stimulus onsets, offsets, and change (Draganova, et al., 2002). Thus, the apparent analogue encoding of the stimulus log-envelope in the auditory cortex more likely represents overlapping cortical responses to stimulus change (i.e., the stimulus envelope derivative). This might also be true of envelope-following responses recorded to low frequency sinusoidally amplitude modulated tones (Maiste & Picton, 1989; Picton, et al., 1987; Rees, et al., 1986) or noise (Liegeois-Chauvel, et al., 2004; Purcell, et al., 2004; Rees, et al., 1986). The derivative of a sinusoid is simply a phase-shifted version of the original sinusoid, so a series of transient responses to the derivative would be difficult to distinguish from an analogue encoding of the sinusoid itself.

The transient response model used in the present study was based on a convolution of the stimulus log-envelope derivative and a P1-N1-P2 complex (the grand average response to the onset of the sentences). This means that the predicted response was the sum of the P1-N1-P2 complex multiplied by every element of the (positive-only) envelope derivative. The model assumes that that a P1-N1-P2 complex is continually elicited by ongoing stimulus change, because a (derivative-weighted) P1-N1-P2 complex is added to the response whenever the derivative is nonzero. The model can thus account for the P1-N1-P2 response to stimulus onset as well as the steady state nature of the response to continuously changing stimuli (e.g., amplitude-modulated noise). The model obviously does not represent the behavior of any individual neuron, but it might be a reasonable representation of the behavior of a large population of neurons. The probability of an individual neural response varies in relation to the rate or acceleration of stimulus change (Heil, 2004; Krishna, 2002). As the probability of a response increases, the number of neurons responding should also increase. Because the evoked response reflects the sum of this activity, it might be reasonably represented as a continuous function of stimulus change (i.e., the stimulus envelope derivative).

The success of the convolution-based model suggests that the P1-N1-P2 complex recorded to a short stimulus change can be treated like the impulse response of a filter. The process of filtering involves the multiplication of a signal (in the frequency domain) with the transfer function of the filter, or (equivalently) the convolution of a signal (in the time domain) with the impulse response of the filter (Hartman, 1998). The impulse response is the result of applying the filter to an impulse, and thus represents the temporal behavior of the system in response to a very basic input. Given that the P1-N1-P2 complex is the typical scalp-recorded response to impulse-like stimuli (e.g., clicks and tone-bursts), it is reasonable to treat it like the impulse response of the auditory cortex.

A system's impulse response defines the frequency response of the system, and also the delay that the system will impose at each frequency. Figure 4 (bottom right) shows the frequency response of the P1-N1-P2 "impulse response." Note that frequencies below 7 Hz are well preserved (the 6 dB-down point), and higher frequencies are attenuated by approximately 15 dB per octave. At frequencies below 7 Hz, the response introduces an average group delay of 209 msec (similar to the latency of P2). The P1-N1-P2 "filter" thus preserves the dominant frequencies of the speech envelope (2 to 7 Hz), and introduces a delay (relative to the envelope derivative) of just over 200 msec.

The P1-N1-P2 complex might be better considered as a series of overlapping impulse responses, because each of the peaks in the "impulse response" likely reflects a separate process (or series of processes) in the cortex. The peaks seem to be generated in different areas (Draganova, et al., 2002; Ponton, et al., 2002; Tiitinen, et al., 1999), and mature at different rates (Ponton, et al., 2002; Wunderlich, et al., 2006). Each of the peaks might also reflect multiple processes. For instance, the N1 may have as many as five distinct subcomponents (Näätänen & Picton, 1987). The components have different refractory periods. With a 500 msec interstimulus interval, the N1 is barely detectable in MEG recordings, and the response is dominated by a large P1 that runs into the P2 (Draganova, et al., 2002; Onitsuka, et al., 2000). For EEG recordings, the N1 and P2 can be detected with interstimulus intervals as short as 200 msec, although the N1 and P2 tend to be similar in size to the P1, and the cortical response looks almost sinusoidal (Picton, et al., 1992). Although the amplitudes of the peaks are sensitive to the interstimulus interval, the latencies of P1, N1, and P2 do not change much as a function of interstimulus interval (Gilley, et al., 2005). If each of the peaks was treated as a separate impulse response, with the appropriate source localization and refractory characteristics, it might be possible to more accurately predict the cortical response.

Efficiency of Response Detection

Although sentence stimuli are ideal for testing individuals wearing hearing aids, the potential clinical value of the cortical response to the speech envelope will ultimately depend on how efficiently and reliably the response can be recorded in individuals. Although all our participants had responses that were significantly correlated with the transient response model, our experimental procedures could not be practically adopted in a clinical setting. To detect a significant response, the response needs to be distinguished from background electrical activity. Simply correlating EEG signals to the speech envelope usually does not give any significant responses because the response is lost in the background EEG activity. EEG activity that is not time-locked to the stimulus can be minimized by averaging responses to multiple stimulus presentations. In the present study, each participant's sentence-specific response was calculated by averaging scalp-recordings from 100 presentations of each of six sentences. The time required for testing could be minimized by using a single sentence as the stimulus, although the generalizability of single-sentence responses has not been established. Envelope-response correlations varied across sentences, as did model-response correlations. This could be because some of the sentences actually elicited stronger responses, or because our prediction of the response was more effective for some of the sentences.

Background electrical activity was also minimized by using multichannel recordings that permitted us to derive source waveforms. Recordings with a large number of channels are impractical for easy clinical use, because of the time required for application and removal of the electrodes. However, the waveforms that correlated best with the envelope and the transient response model were from vertical sources in the auditory cortex, and could likely be efficiently recorded between electrodes placed at the vertex and the nape of the neck. MEG recordings which show tangential sources and eliminate horizontal sources should have better signal to noise ratios for the activity that follows the envelope (e.g., Ahissar, et al., 2001).

It might be possible to improve the efficiency of response detection by improving the model of the response. For a simple amplitude- or frequency-modulated stimulus, the evoked response is expected to occur at the frequency of modulation, and the activity recorded at other frequencies can be treated as noise. The frequency resolution of the analysis can thus be used to separate the response from background electrical activity. For a complex stimulus such as a speech envelope, the evoked response does not occur in a single frequency band. However, if the response can be accurately modeled, the response can be separated from background activity by cross-correlating it with the predicted model. The magnitude of the correlation should be highest when the response and the model are most similar. Therefore, the precision of the model plays a similar role to the frequency resolution of an analysis used to detect a response at a single frequency.

The model that we have used—convolving the onset response with the derivative of the speech envelope—is simplistic in the extreme. This model could be improved in several ways. First, the P1-N1-P2 complex should be modeled as a series of impulse responses with very significant refractory periods. The response to an envelope change occurring 100 msec after the preceding change will be much smaller than the response that follows a 2-sec silence. The refractory effects differ among the separate peaks, with the N1 wave more refractory than the P1 and P2 waves. Second, some threshold of change should be included in the model—a change complex should only occur if the change is above some level of variability. Third, we should likely model simultaneous envelopes for several different frequency bands, using transient responses with latencies specific to those frequencies. The number and variability of the frequency bands will need to be determined. Fourth, it might be better to convolve the impulses with the second derivative of the stimulus envelope instead of the first derivative, because this might bear a closer relationship to the response (Heil, 1997a,b). This approach might allow us to model both onset and offsets. Fifth, it will probably be necessary to consider the response to be more than just a function of the amplitude envelope, because the cortical response to speech is sensitive to the characteristics of the glottal source (e.g., Mäkelä, et al., 2002, 2004; Tiitinen, et al., 2004), and the acoustic change complex can be elicited by changes in periodicity even in the absence of changes to the spectral envelope (Martin & Boothroyd, 1999). Other parameters may be added to the model based on the physiology of cortical processes (cf. May, et al., 1999). These modifications will need to be studied in future research.

Unanswered Questions

Sentences and longer speech segments are appropriate stimuli for eliciting electrophysiologic responses in individuals wearing hearing aids, but we do not know how amplification affects cortical responses to the speech envelope. One would expect the cortical response to vary with the degree to which a hearing aid makes sounds audible, but this remains to be verified (e.g., Tremblay, et al., 2006). It would be useful if cortical responses could be used to predict speech intelligibility.

It is also important to determine the characteristics of the cortical response to the speech envelope in children and young infants. The test might be valuable as a means of validating hearing aid fittings in infants, who cannot provide behavioral responses. But the peaks of the transient response are immature at birth, and the N1 is immature through to adolescence. The maturation of the cortical response to the envelope will therefore need to be extensively studied.

Summary

Sources in the auditory cortex represent the log-envelope of speech at a delay ranging from 152 to 208 msec. The cortex may therefore generate an envelope-following response that synchronizes other analyses of the speech signal. However, this apparent envelope-following response can also be modeled as a series of P1-N1-P2 transient responses convolved with the derivative of the log-envelope. When the temporal relationships between the envelope and its derivative are taken into account, the envelope-following response seems to be predominantly a manifestation of the P2 of the acoustic change complex. Given that onsets and acoustic changes are widely represented by transient responses in the auditory cortex, the envelope-following response could therefore represent a series of overlapping transient responses.

The relationship between the cortical response to the speech envelope and speech intelligibility needs to be determined in both normal-hearing and hearing-impaired adults, children, and infants. The peaks of the cortical response mature at different rates, so the infant and child responses might be very different from the adult response. This needs to be tested. Finally, the effects of amplification on the cortical response to the speech envelope need to be determined.

Acknowledgments

References

Agung, K., Purdy, S. C., McMahon, C. M., et al. (2006). The use of cortical auditory evoked potentials to evaluate neural encoding of speech sounds in adults. *J Am Acad Audiol*, *17*, 559–572.
ExternalResolverBasic │ Bibliographic Links │[Context Link]

Ahissar, E., Nagarajan, S., Ahissar, M., et al. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci U S A*, *98*, 13367–13372.
ExternalResolverBasic │ Bibliographic Links │[Context Link]

Aiken, S. J., & Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiol Neurootol*, *11*, 213–232. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Alonso-Bua, B., Diaz, F., & Ferraces, M. J. (2006). The contribution of AERPs (MMN and LDN) to studying temporal vs. linguistic processing deficits in children with reading difficulties. *Int J Psychophysiol*, *59*, 159–167. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalogr Clin Neurophysiol*, *90*, 229–241. ExternalResolverBasic │ Bibliographic Links │ [Context Link]

Bieser, A., & Muller-Preuss, P. (1996). Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. *Exp Brain Res*, *108*, 273–284. [Context Link]

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Mem Cognit*, *8*, 631–642. [Context Link]

Connolly, J. F., Phillips, N. A., Stewart, S. H., et al. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain Lang*, *43*, 1–18. [Context Link]

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on

neurophysiological responses to spoken sentences. *Brain Lang*, *39*, 302–318.   ExternalResolverBasic | Bibliographic Links | [Context Link]

D'Arcy, R. C., Connolly, J. F., Service, E, et al. (2004). Separating phonological and semantic processing in auditory sentence processing: a high resolution event-related potential study. *Hum Brain Mapp*, *22*, 40–51. [Context Link]

Dimitrijevic, A., John, M. S., & Picton, T. W. (2004). Auditory steady-state responses and word recognition scores in normal-hearing and hearing-impaired adults. *Ear Hear*, *25*, 68–84.   Ovid Full Text | ExternalResolverBasic | Bibliographic Links | [Context Link]

Dobie, R. A., & Wilson, M. J. (1984). Short-latency auditory responses obtained by cross correlation. *J Acoust Soc Am*, *76*, 1411–1421.   ExternalResolverBasic | Bibliographic Links | [Context Link]

Draganova, R., Ross, B., Borgmann, C., et al. (2002). Auditory cortical response patterns to multiple rhythms of AM sound. *Ear Hear*, *23*, 254–265.   Ovid Full Text | ExternalResolverBasic | Bibliographic Links | [Context Link]

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am*, *95*, 1053–1064. [Context Link]

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of reducing slow temporal modulations on speech reception. *J Acoust Soc Am*, *95*, 2670–2680. [Context Link]

Eggermont, J. J. (1997). Firing rate and firing synchrony distinguish dynamic from steady state sound. *Neuroreport*, *8*, 2709–2713.   Ovid Full Text | ExternalResolverBasic | Bibliographic Links | [Context Link]

Eggermont, J. J., & Ponton, C. W. (2002). The neurophysiology of auditory perception: from single-units to evoked potentials. *Audiol Neurootol*, *7*, 71–99. [Context Link]

Eulitz, C., Diesch, E., Pantev, C., et al. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *J Neurosci*, *15*, 2748–2755. [Context Link]

Gilley, P. M., Sharma, A., Dorman, M., et al. (2005). Developmental changes in refractoriness of the cortical auditory evoked potential. *Clin Neurophysiol*, *116*, 648–657.   ExternalResolverBasic | Bibliographic Links | [Context Link]

Hartman, W. M. (1998). *Signals, sound and sensation*. New York: Springer-Verlag. [Context Link]

Heil, P. (1997a). Auditory cortical onset responses revisited. I. First-spike timing. *J Neurophysiol*, *77*, 2616–2641. [Context Link]

Heil, P. (1997b). Auditory cortical onset responses revisited. II. Response strength. *J Neurophysiol*, *77*, 2642–2660. [Context Link]

Heil, P. (2004). First-spike latency of auditory neurons revisited. *Curr Opin Neurobiol*, *14*, 461–467. [Context Link]

Herdman, A. T., Lins, O., Van Roon, P., et al. (2002). Intracerebral sources of human auditory steady-state

responses. *Brain Topogr*, 15, 69–86. ExternalResolverBasic | Bibliographic Links | [Context Link]

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. *J Clin Neurophysiol, 19*, 113–124. Ovid Full Text | ExternalResolverBasic | Bibliographic Links | [Context Link]

Ilvonen, T., Kujala, T., Kozou, H., et al. (2004). The processing of speech and non-speech sounds in aphasic patients as reflected by the mismatch negativity (MMN). *Neurosci Lett*, 366, 235–240. ExternalResolverBasic | Bibliographic Links | [Context Link]

Josse, G., & Tzourio-Mazoyer, N. (2004). Hemispheric specialization for language. *Brain Res Brain Res Rev, 44*, 1–12. [Context Link]

Kelly, A. S., Purdy, S. C., & Thorne, P. R. (2005). Electrophysiological and speech perception measures of auditory processing in experienced adult cochlear implant users. *Clin Neurophysiol, 116*, 1235–1246. ExternalResolverBasic | Bibliographic Links | [Context Link]

Korczak, P. A., Kurtzberg, D., & Stapells, D. R. (2005). Effects of sensorineural hearing loss and personal hearing aids on cortical event-related potential and behavioral measures of speech-sound processing. *Ear Hear, 26*, 165–185. Ovid Full Text | ExternalResolverBasic | Bibliographic Links | [Context Link]

Kraus, N., Bradlow, A. R., Cheatham, M. A., et al. (2000). Consequences of neural asynchrony: a case of auditory neuropathy. *J Assoc Res Otolaryngol, 1*, 33–45. ExternalResolverBasic | Bibliographic Links | [Context Link]

Krishna, B. S. (2002). A unified mechanism for spontaneous-rate and first-spike timing in the auditory nerve. *J Comput Neurosci, 13*, 71–91. [Context Link]

Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res, 166*, 192–201. [Context Link]

Krishnan, A., Xu, Y., Gandour, J. T., et al. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res, 189*, 1–12. [Context Link]

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203–205. ExternalResolverBasic | Bibliographic Links | [Context Link]

Liang, L., Lu, T., & Wang, X. (2002). Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. *J Neurophysiol, 87*, 2237–2261. ExternalResolverBasic | Bibliographic Links | [Context Link]

Liégeois-Chauvel, C., Lorenzi, C., Trebuchon, A., et al. (2004). Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex, 14*, 731–740. ExternalResolverBasic | Bibliographic Links | [Context Link]

Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci, 4*, 1131–1138. ExternalResolverBasic | Bibliographic Links | [Context Link]

Maiste, A., & Picton, T. (1989). Human auditory evoked potentials to frequency-modulated tones. *Ear Hear*,

10, 153–160. Ovid Full Text | ExternalResolverBasic | Bibliographic Links | [Context Link]

Mäkelä, A.M., Alku, P., Mäkinen, V., et al. (2002). Human cortical dynamics determined by speech fundamental frequency. *Neuroimage, 17*, 1300–1305. ExternalResolverBasic | Bibliographic Links | [Context Link]

Mäkelä, A.M., Alku, P., Mäkinen, V., et al. (2004). Glides in speech fundamental frequency are reflected in the auditory N1m response. *Neuroreport*, *15*, 1205–1208. ExternalResolverBasic | Bibliographic Links | [Context Link]

Mäkelä, A.M., Alku, P., & Tiitinen, H. (2003). The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. *Neurosci Lett*, *353*, 111–114. ExternalResolverBasic | Bibliographic Links | [Context Link]

Martin, B. A., & Boothroyd, A. (1999). Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear Hear*, *20*, 33–44. Ovid Full Text | ExternalResolverBasic | Bibliographic Links | [Context Link]

Martin, B. A., Sigal, A., Kurtzberg, D., et al. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *J Acoust Soc Am*, *101*, 1585–1599. ExternalResolverBasic | Bibliographic Links | [Context Link]

Martin, B. A., Tremblay, K. L., & Stapells, D.R. (2007). Principles and applications of cortical evoked potentials. In R. F. Burkard, M. Don, & J. J. Eggermont (Eds.). *Auditory evoked potentials: basic principles and clinical applications.* (pp. 482–507). Baltimore: Lippincott, Williams and Wilkins. [Context Link]

May, P., Tiitinen, H., Ilmoniemi, R. J., et al. (1999). Frequency change detection in human auditory cortex. *J Comput Neurosci*, *6*, 99–120. ExternalResolverBasic | Bibliographic Links | [Context Link]

Näätänen, R., & Picton, T. W. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425. ExternalResolverBasic | Bibliographic Links | [Context Link]

Oates, P. A., Kurtzberg, D., & Stapells, D. R. (2002). Effects of sensorineural hearing loss on cortical event-related potential and behavioral measures of speech-sound processing. *Ear Hear*, *23*, 399–415. Ovid Full Text | ExternalResolverBasic | Bibliographic Links | [Context Link]

Onitsuka, T., Ninomiya, H., Sato, E., et al. (2000). The effect of interstimulus intervals and between-block rests on the auditory evoked potential and magnetic field: is the auditory P50 in humans an overlapping potential? *Clin Neurophysiol*, *111*, 237–245. [Context Link]

Ostroff, J. M., Martin, B. A., & Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear Hear*, *19*, 290–297. Ovid Full Text | ExternalResolverBasic | Bibliographic Links | [Context Link]

Oygarden, J. (2005). Experiments with steady-state evoked responses from speech and music. Paper presented at the XIX Meeting of the International Evoked Response Audiometry Study Group, Havana, Cuba, June 12–16. [Context Link]

Phillips, D. P., Hall, S. E., & Boehnke, S. E. (2002). Central auditory onset responses, and temporal

asymmetries in auditory perception. *Hear Res*, *167*, 192–205. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Picton, T. W., Kellett, A. J. C., & Champagne, S. C. (1992). Auditory evoked potentials recorded using maximum length sequences. *Electroencephalogr Clin Neurophysiol*, *84*, 90–100. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Picton, T. W., Skinner, C. R., Champagne, S. C., et al. (1987). Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *J Acoust Soc Am*, *82*, 165–178. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Picton, T. W., van Roon, P., Armilio, M. L., et al. (2000). The correction of ocular artifacts: a topographic perspective. *Clin Neurophysiol*, *111*, 53–65. [Context Link]

Plyler, P., & Krishnan, R. (2001). Human frequency-following responses: representation of second formant transitions in normal and hearing-impaired listeners. *J Am Acad Audiol*, *12*, 523–533. [Context Link]

Ponton, C., Eggermont, J. J., Khosla, D., et al. (2002). Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. *Clin Neurophysiol*, *113*, 407–420. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Price, C. J., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: where is it? *Trends Cogn Sci*, *9*, 271–276. [Context Link]

Purcell, D. W., John, M. S., Schneider, B. A., et al. (2004). Human temporal auditory acuity as assessed by envelope following responses. *J Acoust Soc Am*, *116*, 3581–3593. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Purdy, S. C., Kelly, A. S., & Davies, M. G. (2002). Auditory brainstem response, middle latency response, and late cortical evoked potentials in children with learning disabilities. *J Am Acad Audiol*, *13*, 367–382. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Rance, G., Cone-Wesson, B., Wunderlich, J., et al. (2002). Speech perception and cortical event related potentials in children with auditory neuropathy. *Ear Hear*, *23*, 239–253. Ovid Full Text │ ExternalResolverBasic │ Bibliographic Links │[Context Link]

Rees, A., Green, G. G., & Kay, R. H. (1986). Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear Res*, *23*, 123–133. ExternalResolverBasic │ Bibliographic Links │ [Context Link]

Russo, N., Nicol, T., Musacchia, G., et al. (2004). Brainstem responses to speech syllables. *Clin Neurophysiol*, *115*, 2021–2030. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Schreiner, C. E., & Urbas, J. V. (1986). Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). *Hear Res*, *21*, 227–241. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Schulte-Korne, G., Deimel, W., Bartling, J., et al. (1998). Auditory processing and dyslexia: evidence for a specific speech processing deficit. *Neuroreport*, *9*, 337–340. Ovid Full Text │ ExternalResolverBasic │ Bibliographic Links │[Context Link]

Scollie, S. D., & Seewald, R. C. (2002). Evaluation of electroacoustic test signals. I. Comparison with amplified speech. *Ear Hear*, *23*, 477–487. Ovid Full Text │ ExternalResolverBasic │ Bibliographic Links │ [Context Link]

Shannon, R. V., Zeng, F. G., Kamath, V., et al. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Shtyrov, Y., Pihko, E., & Pulvermuller, F. (2005). Determinants of dominance: is language laterality explained by physical or linguistic features of speech? *Neuroimage*, *27*, 37–47. [Context Link]

Smith, J. C., Marsh, J. T., & Brown, W. S. (1975). Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol*, *39*, 465–472. [Context Link]

Stone, M. A., & Moore, B. C. J. (1992). Syllabic compression: effective compression ratios for signals modulated at different rates. *Br J Audiol*, *26*, 351–361. ExternalResolverBasic │ Bibliographic Links │ [Context Link]

Studebaker, G. A., & Sherbecoe, R. L. (2002). Intensity-importance functions for bandlimited monosyllabic words. *J Acoust Soc Am*, *111*, 1422–1436. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Res Brain Res Rev*, *43*, 231–246. [Context Link]

Tiitinen, H., Mäkelä, A.M., Mäkinen, V., et al. (2004). Periodic glottal excitation and formant frequencies in the perception of vowels. *Neurol Clin Neurophysiol*, *2004*, 103–107. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Tiitinen, H., Sivonen, P., Alku, P., et al. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. *Brain Res Cogn Brain Res*, *8*, 355–363. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Tremblay, K. L., Billings, C. J., Friesen, L. M., et al. (2006a). Neural representation of amplified speech sounds. *Ear Hear*, *27*, 93–103. Ovid Full Text │ ExternalResolverBasic │ Bibliographic Links │[Context Link]

Tremblay, K. L., Billings, C., & Rohila, N. (2004). Speech evoked cortical potentials: effects of age and stimulus presentation rate. *J Am Acad Audiol*, *15*, 226–237. ExternalResolverBasic │ Bibliographic Links │ [Context Link]

Tremblay, K. L., Friesen, L., Martin, B. A., et al. (2003a). Test-retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear Hear*, *24*, 225–232. [Context Link]

Tremblay, K. L., Kalstein, L., Billings, C. J., et al. (2006b). The neural representation of consonant-vowel transitions in adults who wear hearing aids. *Trends Amplif*, *10*, 155–162. ExternalResolverBasic │ Bibliographic Links │[Context Link]

Tremblay, K. L., Piskosz, M., & Souza, P. (2002). Aging alters the neural representation of speech cues. *Neuroreport*, *13*, 1865–1870. Ovid Full Text │ ExternalResolverBasic │ Bibliographic Links │[Context Link]

Tremblay, K. L., Piskosz, M., & Souza, P. (2003b). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clin Neurophysiol*, *114*, 1332–1343. [Context Link]

Uwer, R., Albrecht, R., & von Suchodoletz, W. (2002). Automatic processing of tones and speech stimuli in children with specific language impairment. *Dev Med Child Neurol*, *44*, 527–532. [Context Link]

van der Horst, R., Leeuw, A. R., & Dreschler, W. A. (1999). Importance of temporal-envelope cues in consonant recognition. *J Acoust Soc Am*, *105*, 1801–1809. ExternalResolverBasic | Bibliographic Links | [Context Link]

Van Tasell, D. J., Soli, S. D., Kirby, V. M., et al. (1987). Speech waveform envelope cues for consonant recognition. *J Acoust Soc Am*, *82*, 1152–1161. ExternalResolverBasic | Bibliographic Links | [Context Link]

Wang, X., Lu, T., Snider, R. K., et al. (2005). Sustained firing in auditory cortex evoked by preferred stimuli. *Nature*, *435*, 341–346. ExternalResolverBasic | Bibliographic Links | [Context Link]

Wunderlich, J. L., Cone-Wesson, B. K., & Shepherd, R. (2006). Maturation of the cortical auditory evoked potential in infants and young children. *Hear Res*, *212*, 185–202. ExternalResolverBasic | Bibliographic Links | [Context Link]

Yamada, Y., & Neville, H. J. (2007). An ERP study of syntactic processing in English and nonsense sentences. *Brain Res*, *1130*, 167–180. [Context Link]

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., et al. (2006). Top-down and bottom-up processes in speech comprehension. *Neuroimage*, *32*, 1826–1836. [Context Link]

IMAGE GALLERY

Select All   Export Selected to PowerPoint
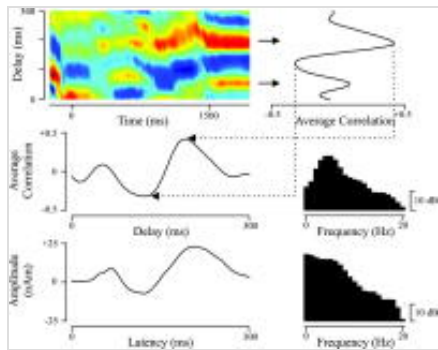

Fig. 1

Fig. 2

Fig. 3

TABLE 1. Amplitudes and latencies of grand average onset response to all sentences (vertical sources)

$$y(t) = \sum_{k=0}^{N} x(k)h(t-k)$$

☐ Fig. 4

☐ Table 1

☐ Equation (Uncited)

☐ Fig. 5

☐ Fig. 6

☐ Fig. 7

☐ Fig. 8

☐ Table 2

☐ Fig. 9

Back to Top

About Us | Contact Us | Terms of Use