

Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex

Kirill V. Nourski,¹ Richard A. Reale,^{1,3} Hiroyuki Oya,¹ Hiroto Kawasaki,¹ Christopher K. Kovach,^{2,4} Haiming Chen,¹ Matthew A. Howard III,¹ and John F. Brugge^{1,3}

Departments of ¹Neurosurgery and ²Neurology, The University of Iowa, Iowa City, Iowa 52242, ³Department of Psychology, University of Wisconsin–Madison, Madison, Wisconsin 53706, and ⁴Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125

Speech comprehension relies on temporal cues contained in the speech envelope, and the auditory cortex has been implicated as playing a critical role in encoding this temporal information. We investigated auditory cortical responses to speech stimuli in subjects undergoing invasive electrophysiological monitoring for pharmacologically refractory epilepsy. Recordings were made from multicontact electrodes implanted in Heschl's gyrus (HG). Speech sentences, time compressed from 0.75 to 0.20 of natural speaking rate, elicited average evoked potentials (AEPs) and increases in event-related band power (ERBP) of cortical high-frequency (70–250 Hz) activity. Cortex of posteromedial HG, the presumed core of human auditory cortex, represented the envelope of speech stimuli in the AEP and ERBP. Envelope following in ERBP, but not in AEP, was evident in both language-dominant and -nondominant hemispheres for relatively high degrees of compression where speech was not comprehensible. Compared to posteromedial HG, responses from anterolateral HG—an auditory belt field—exhibited longer latencies, lower amplitudes, and little or no time locking to the speech envelope. The ability of the core auditory cortex to follow the temporal speech envelope over a wide range of speaking rates leads us to conclude that such capacity in itself is not a limiting factor for speech comprehension.

Introduction

The temporal envelope of human speech reflects amplitude fluctuations ranging from ~2 Hz to 50 Hz, which correspond to phonemic and syllabic transitions critically important for comprehension (Rosen, 1992). Speech recognition can be achieved when the spectral information is severely limited but temporal envelope cues are preserved (Shannon et al., 1995). Comprehension of speech auditory chimeras (in which the envelope of one stimulus is used to modulate the fine structure of another) is based primarily on envelope cues (Smith et al., 2002). Distorting the speech envelope by temporal smearing (Drullman et al., 1994) or compression (Ahissar and Ahissar, 2005) impairs comprehension.

Understanding how and where speech envelope information is represented within human auditory cortex continues to be a major challenge (Luo and Poeppel, 2007). Ahissar et al. (2001), using magnetoencephalography (MEG), observed that degraded comprehension of time-compressed speech correlated with a decline in temporal synchrony between auditory cortical responses and the speech envelope. They placed this processing mechanism

“approximately on Heschl's gyrus” and concluded that temporal locking of activity in this cortical area to the speech envelope was a prerequisite for comprehension. The modal frequency of the most compressed speech signal used by Ahissar et al. (2001) was ~14 Hz, which, in humans, is well within the limits of phase locking to the envelope of sinusoidal amplitude-modulated tones and noise (Kuwada et al., 1986; Rees et al., 1986; Roß et al., 2000; Liégeois-Chauvel et al., 2004; Nourski et al., 2009). These findings support an alternative hypothesis, that the auditory cortex of Heschl's gyrus (HG) can temporally encode the speech envelope even at high modulation rates beyond speech comprehension. We tested this hypothesis by recording directly from HG, in human neurosurgical subjects, activity evoked by speech stimuli that were essentially identical to those used by Ahissar et al. (2001). We were able to accurately localize the evoked activity within the gyrus anatomically and physiologically.

Using the intracortical recording approach, primary and primary-like auditory cortices (the auditory core) have been localized to posteromedial HG (Liégeois-Chauvel et al., 1991; Howard et al., 1996, 2000; Brugge et al., 2008). Average evoked potentials (AEPs) (Donchin and Lindsley, 1969) recorded there have a relatively short latency and feature phase-locked responses to periodic stimuli. These properties distinguish the core from an auditory field on anterolateral HG, which exhibits AEPs having longer latency with little evidence of phase locking to the stimulus. This laterally positioned field has been interpreted as an auditory cortical belt system.

High-frequency cortical activity (above ~70 Hz) has been shown to be a prominent component of auditory cortical responses

Received June 28, 2009; revised Oct. 7, 2009; accepted Nov. 7, 2009.

This study was supported by National Institute on Deafness and Other Communication Disorders Grant R01-DC04290, General Clinical Research Centers Program Grant M01-RR-59, the Hoover Fund, and the Carver Trust. We thank Mitchell Steinschneider, Christopher Turner, and Paul Poon for advice and comments, Ehud Ahissar for providing experimental stimuli, Chandan Reddy and Fangxiang Chen for help during data collection, and Carol Dizack for graphic artwork.

Correspondence should be addressed to Kirill V. Nourski, Department of Neurosurgery, The University of Iowa, 200 Hawkins Drive, 1825 JPP, Iowa City, IA 52242. E-mail: kirill-nourski@uiowa.edu.

DOI:10.1523/JNEUROSCI.3065-09.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/2915564-11\$15.00/0

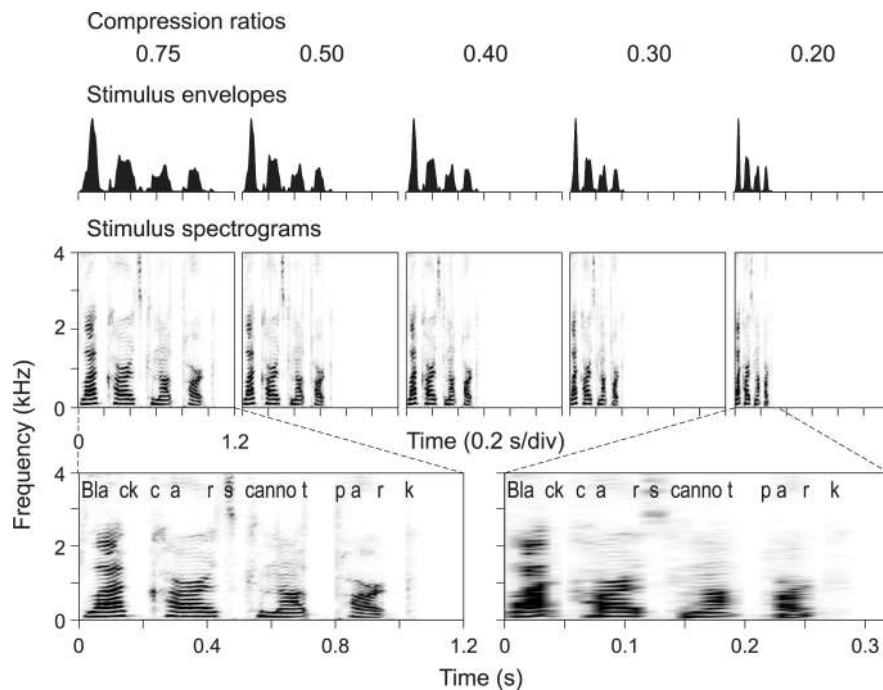


Figure 1. Temporal envelopes (top row) and spectrograms (middle and bottom row) of time-compressed stimuli (speech sentence “Black cars cannot park”) used in the experiments. In the bottom row, spectrograms of stimuli compressed to ratios of 0.75 (least compressed) and 0.20 (most compressed) are replotted to illustrate preservation of spectral content across compression ratios.

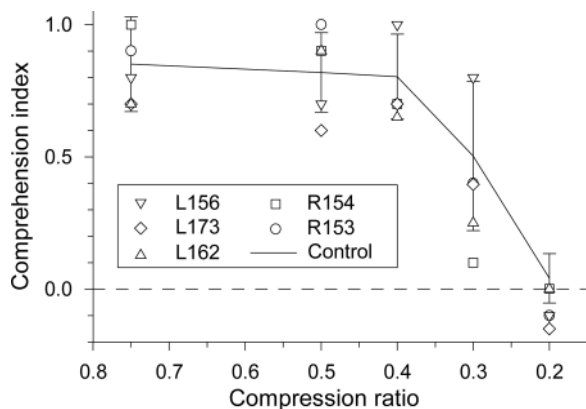


Figure 2. Comprehension of time-compressed speech sentences by the neurosurgical subject patients (symbols) and in a control group of healthy subjects ($n = 20$) (mean \pm SD; lines with error bars). The subjects’ performance in the psychophysical task was measured as the comprehension index (see Materials and Methods for details). Dashed line indicates chance level.

in human and monkey (Crone et al., 2001; Steinschneider et al., 2008). In this study, we used time–frequency analysis of single-trial response waveforms to capture event-related band power (ERBP) of the electrocorticogram (ECoG) within the frequency range of 70–250 Hz. We explored the relationship between stimulus temporal envelope and cortical activity, measured as the AEP as well as ERBP, at multiple recording sites within HG of both the language-dominant and -nondominant hemispheres.

Materials and Methods

Experimental subjects. The six subjects (two males, four females; 22–45 years old) that participated in this study were neurosurgical patients diagnosed with pharmacologically refractory epilepsy and were undergoing chronic invasive electroencephalography monitoring to identify a

seizure focus before surgical treatment. Written informed consent was obtained from each subject. Research protocols were approved by The University of Iowa Human Subjects Review Board.

All six subjects were right handed, and one (L162) had mixed language dominance, while the five others had left hemisphere language dominance, as determined by Wada test results. In three of the six subjects studied (L156, L162, L173), the electrodes were implanted on the left side, while in three other subjects (R152, R153, R154), recordings were made from the right hemisphere. All subjects underwent audiometric and neuropsychological evaluation before the study, and none were found to have hearing or cognitive deficits that would impact the findings presented in this study. All subjects were native English speakers. Analysis of intracranial recordings indicated that HG was not involved in the generation of epileptic activity in the subjects.

Stimulus presentation. Experimental stimuli were speech sentences, digitized at a sampling rate of 24,414 Hz. The stimuli were time compressed to ratios 0.75, 0.50, 0.40, 0.30, and 0.20 of the natural speaking rate (Fig. 1) using an algorithm that preserved the spectral content of the stimuli, as implemented in Sound Designer II software.

Evaluation of comprehension of time-compressed speech sentences was performed following the approach of Ahissar et al. (2001).

The psychophysical experiment was performed in five of six subjects (all except R152) and in a control group of 20 healthy volunteers (13 males, 7 females, 19–35 years old, all native English speakers). The following set of 10 sentences was used (T indicates a true statement, F indicates a false statement): (1) Black cars can all park: T. (2) Black dogs can all bark: T. (3) Black cars cannot bark: T. (4) Black dogs cannot park: T. (5) Playing cards cannot park: T. (6) Black cars cannot park: F. (7) Black dogs cannot bark: F. (8) Black cars can all bark: F. (9) Black dogs can all park: F. (10) Playing cards can all park: F.

Each sentence was presented at five compression ratios in a random order, and each sentence was presented twice at each compression ratio, thus yielding a total of 100 trials in the psychophysical experiment. The subjects were instructed to respond to the sentences by pressing one of three buttons, corresponding to “True,” “False,” or “I don’t know.” Comprehension was quantified using a comprehension index (CI) (Ahissar et al., 2001; Ahissar and Ahissar, 2005), calculated as follows: $CI = (N_{correct} - N_{incorrect}) / N_{total}$, where $N_{correct}$ is the number of correct responses (i.e., “True” statements identified as “True,” and “False” statements identified as “False”); $N_{incorrect}$ is the number of incorrect responses (i.e., “True” statements identified as “False,” and “False” statements identified as “True”); and N_{total} is the total number of trials, including correct responses, incorrect responses, and trials to which the subjects responded with “I don’t know.”

The electrophysiological experiment, performed in the six subjects, used a set of six time-compressed speech stimuli. Five of the stimuli were time-compressed versions of the sentence “Black cars cannot park,” presented at compression ratios of 0.75, 0.50, 0.40, 0.30, and 0.20 (Fig. 1). The sixth stimulus, “Black dogs can all bark,” presented at a compression ratio of 0.75, was used as a target in an oddball detection task to maintain the subject in an alert state. The subjects were instructed to press a button whenever the oddball stimulus was detected. The output of the response box was monitored with an oscilloscope during the recording sessions. The sounds were delivered binaurally via insert earphones (ER4B, Etymotic Research) mounted in subject-specific custom made earmolds. Each stimulus was presented 50 times in random order at a comfortable level (45–55 dB above hearing threshold). The duration of the speech stimuli ranged from 0.29 to 1.05 s (at compression ratios of 0.20 and 0.75,

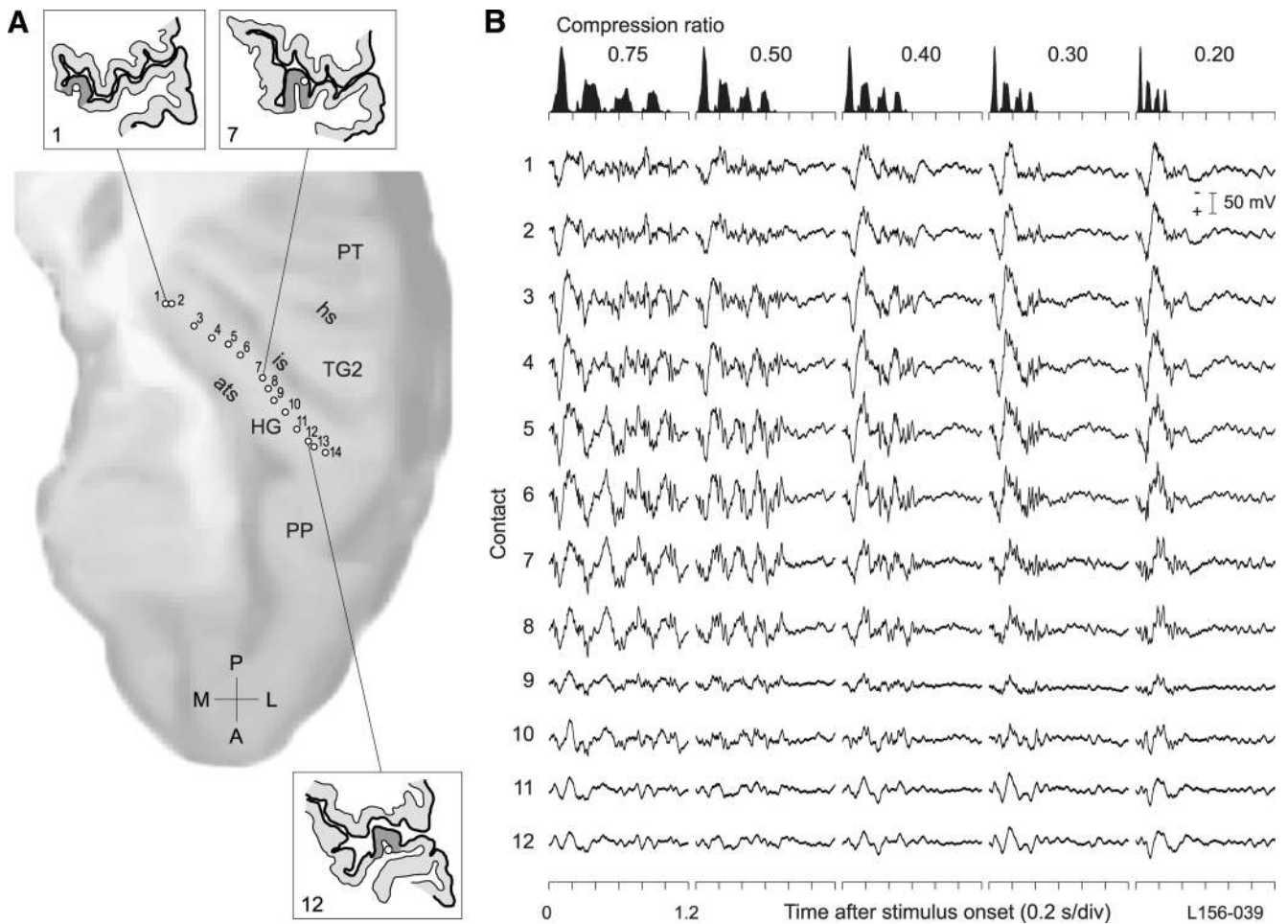


Figure 3. AEPs recorded from left HG in a representative subject. **A**, MRI surface rendering of the superior temporal plane showing location of the micro recording contacts. Macrocontacts used for recording of clinical data are not shown. Insets, Tracings of MRI cross sections showing the location of the recording contacts (open circles) within the gray matter at three representative locations. Dark gray shading denotes the estimated extent of the HG. **B**, AEP waveforms across compression ratios (left to right: moderate to severe compression) and the length of HG (top to bottom: posteromedial to anterolateral). Temporal envelopes of the speech stimuli are shown in the top panels. Recordings from contacts 13 and 14, which were located outside HG gray matter, are not shown. Negative voltage is plotted upward. HG, Heschl's gyrus (anterior transverse gyrus); TG2: second transverse gyrus; PP, planum polare; PT, planum temporale; ats, anterior transverse sulcus; is, intermediate sulcus; hs, Heschl's sulcus.

respectively). The interval between stimulus onsets was fixed at 3 s. Stimulus delivery and data acquisition were controlled by a TDT RX5 or RZ2 processor.

Response recording. Details of electrode implantation have been described previously (Howard et al., 1996, 2000; Brugge et al., 2008; Reddy et al., 2009). In brief, custom-designed hybrid depth electrode (HDE) arrays were implanted stereotactically into HG, along its anterolateral to posteromedial axis. HDEs included six platinum macrocontacts, spaced 10 mm apart, which were used to record clinical data. Fourteen platinum microcontacts (diameter 40 μ m, impedance 0.08–0.7 M Ω) were distributed at 2–4 mm intervals between the macrocontacts and were used to record intracortical ECoG. The reference for the microcontacts was either a subgaleal contact or one of the two most lateral macrocontacts near the lateral surface of the superior temporal gyrus. Reference electrodes, including those near the lateral surface of the superior temporal gyrus, were relatively inactive compared to the large amplitude activity recorded from more medial portions of HG. Recording electrodes remained in place for ~2–3 weeks under the direction of the clinical epileptologists.

Each subject underwent whole-brain MRI and CT scanning before electrode implantation. To locate recording contacts on the HDEs, high-resolution T1-weighted structural MRIs (in-plane resolution 0.78 \times 0.78 \times 1.0 mm) were obtained both before and after electrode implantation. Preimplantation and postimplantation MRIs were coregistered using a three-dimensional rigid fusion algorithm (Analyze version 8.1 software,

Mayo Clinic). Coordinates for each electrode contact obtained from postimplantation MRI volumes were transferred to preimplantation MRI volumes. Serial MR cross-sectional images containing the recording contacts were obtained perpendicular to the trajectory of the HDEs. The coordinates of the electrode shaft were determined using custom-designed software written in the MATLAB programming environment.

ECoG signals were recorded simultaneously from the intracranial HDE contacts, amplified, filtered (1.6–6000 Hz bandpass, 12 dB/octave rolloff), digitized at a sampling rate of 12,207 Hz, and stored for subsequent offline analysis.

Data analysis. Envelopes of the speech stimuli were obtained by calculating the magnitude of the Hilbert transform of the speech signal waveform and low-pass filtering at 50 Hz using a fourth order Butterworth filter. ECoG obtained from each recording site were down-sampled to a sampling rate of 4069 Hz for computational efficiency. Trials that might be contaminated with noise (movement artifacts or electrical interference), and whose maximum amplitude deviated >2.5 SD above the mean, were excluded from the analysis. Data analysis was performed using custom software (MATLAB version 7.7.0).

In the time domain, stimulus-related phase-locked activity in the ECoG was characterized by the AEP. The AEP estimates the most likely response waveform that would result from a single stimulus presentation, if stationary random noise was removed from the recorded voltage measurements. The rationale for this simple averaging approach is the explicit model that this response waveform (i.e., the AEP) is invariant (in

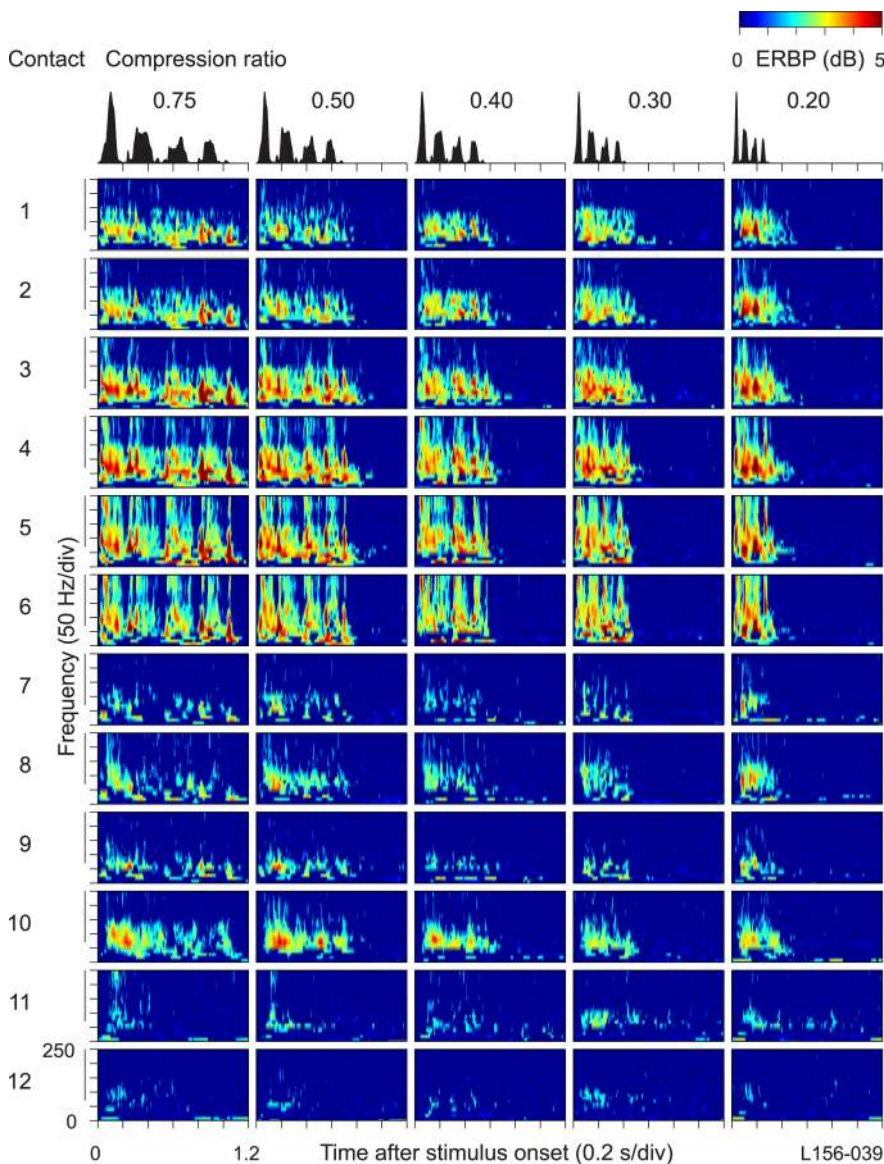


Figure 4. ERBP analysis of recordings from left HG (same subject as in Fig. 3) across compression ratios (left to right: moderate to severe compression) and the length of HG (top to bottom: posteromedial to anterolateral). Temporal envelopes of the speech stimuli are shown in the top panels. Recordings from contacts 13 and 14, which were located outside HG gray matter, are not shown. ERBP was measured in dB relative to power in a reference period of 0.2–0.1 s before stimulus onset. Vertical lines indicate the frequency band (70–250 Hz) in which ERBP envelopes were calculated.

amplitude values and onset latency) for all presentations of an identical stimulus. Therefore, in this homogeneous population of response waveforms, the AEP can be said to be “phase locked” to the stimulus. An alternative model is that response waveforms constitute an inhomogeneous set and are not invariant across identical stimulus trials. Simple averaging is not appropriate to estimate a most likely response waveform under this model and some form of single-trial analysis must be used (Woody, 1967; Crone et al., 1998; Knuth et al., 2006). This may result because the assumption of stationary independent noise is insufficient to characterize the physiological recordings and/or to systematic variability in response waveforms due to unobserved covariates (e.g., adaptation, habituation, learning, etc.). In this single-trial analysis approach, the response waveform is said to be “time locked” to the stimulus given an operational definition of a response-time window.

Time-domain waveform averaging minimizes the contribution of time-but non-phase-locked (NPL) activity that may be important components of the neural activity evoked by speech. This is especially relevant for higher frequencies in the ECoG (Crone et al., 1998; Steinschneider et

al., 2008). Thus, in addition to computing the AEP, the power in selected frequency bands in the ECoG signal was computed to obtain measures of the time-locked but not phase-locked response. This ERBP reflects the increase or decrease in total power in a given frequency band with reference to the ongoing background ECoG (Crone et al., 1998; Pfurtscheller and Lopes da Silva, 1999). Thus the ERBP will include both phase-locked (often termed “evoked”) power (Pantev, 1995) as well as non-phase-locked, yet time-locked (often termed “induced”) power (Kalcher and Pfurtscheller, 1995; Pantev, 1995; Crone et al., 2001).

Time-frequency analysis of the ECoG was performed using wavelet transforms based on complex Morlet wavelets following the approach of Oya et al. (2002). Center frequencies ranged from 10 to 250 Hz in 10 Hz steps, and the constant ratio was defined as $2\pi f_0\sigma = 7$, where f_0 is the center frequency and σ defines the wavelet width. Power measurements were done on a trial-by-trial basis and then averaged across trials. To quantify power changes as ERBP, mean power values were calculated at each center frequency within a reference period of 300 ms before the onset of the stimuli. ERBP values were then calculated at each center frequency and each time point in dB relative to mean power over the reference period. An advantage of such an approach is that power is normalized independently in each frequency band, thus ensuring that the $1/f$ statistical behavior of the ECoG power spectra does not impact the analysis.

While most time-frequency analyses presented in this study measured total power, we also estimated NPL cortical activity in a limited dataset. In this estimation procedure, the contribution of phase-locked response components was minimized using the approach of Crone et al. (2001), by subtracting the AEP from each individual trial waveform before the wavelet transformation.

ERBP envelopes were calculated as log-transformed power changes, normalized, and averaged, over the range of frequencies between 70 and 130 Hz in subject R153, and between 70 and 250 Hz in the other five subjects. The range used for data collected from subject R153 differed from the others due to noise contamination of unknown origin that affected the recorded ECoG at frequencies of >130 Hz.

Representation of the temporal stimulus envelope in the cortical activity was quantified in the time domain using cross-correlation analysis (Bieser and Müller-Preuss, 1996; Abrams et al., 2008) and, in the frequency domain, as modal frequency matching (Ahissar and Ahissar, 2005). Peaks of cross-correlograms were found between lags of 0 and 150 ms. Ninety-five percent confidence intervals of the cross-correlation peaks were calculated based on 1000 bootstrapped samples.

Power spectra of time-compressed speech stimulus envelopes, ECoG single trial waveforms, and ERBP envelopes were estimated using Thomson multitaper approach (Thomson, 1982) as implemented in MATLAB version 7.7.0. The spectrum estimation algorithm was applied with a time-bandwidth product of 1.5 following removal of linear trend. The power spectra of the stimulus envelopes were characterized by their modal frequencies, which ranged from 3.7 to 14 Hz (at compression ratios of 0.75 and 0.20, respectively) (supplemental Fig. 1, available at www.jneurosci.org as supplemental material). Modal frequencies of

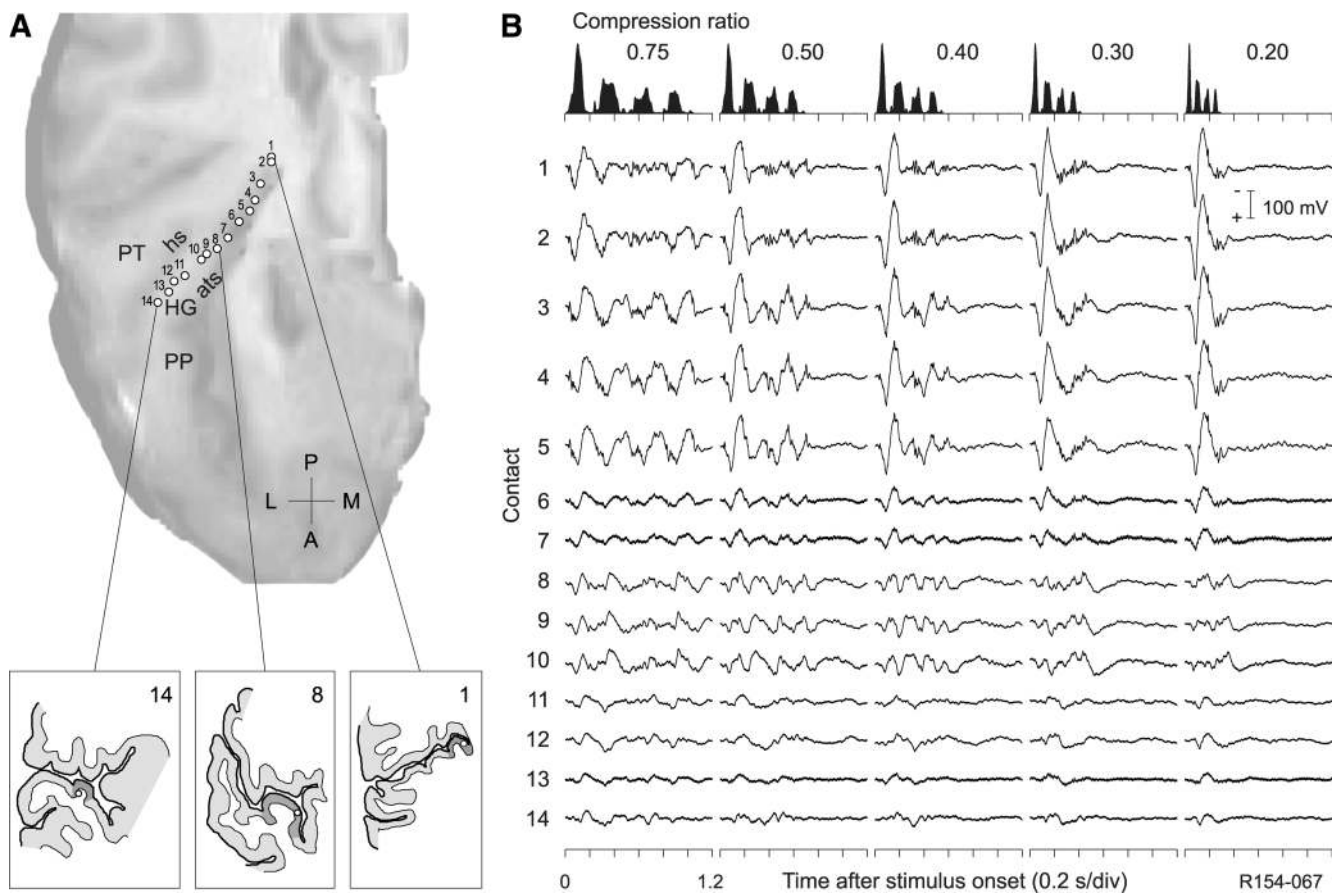


Figure 5. AEPs recorded from right HG in a representative subject. See legend of Figure 3 for details.

ECoG averaged power spectra and ERBP spectra were defined as maximal spectral peaks at frequencies above the reciprocal of the stimulus duration. Peaks below this frequency were ignored because they were likely to represent artifacts of zero-padding and detrending in the context of a DC offset in the ERBP.

Stimulus–response frequency matching was evaluated for the raw ECoG signal as well as for ERBP envelope from their power spectra. In the former case, frequency matching was measured as the difference between modal frequency of the stimulus envelope and the local maximum of the averaged spectrum of ECoG, and in the latter case, as the difference between the modal frequency of the stimulus envelope and the local maximum of the ERBP envelope.

Results

Comprehension of time-compressed speech sentences

Intelligibility of time-compressed speech sentences was evaluated in a psychophysical experiment, the results of which are presented in Figure 2. At compression ratios of 0.75, 0.50, and 0.40, comprehension index values were relatively high (≥ 0.6) in all tested subjects, corresponding to correct identification of at least 80% of the sentences. This indicates that speech sentences presented at these compression ratios were intelligible. At compression ratio of 0.30, speech comprehension deteriorated, and comprehension of sentences compressed to 0.20 of the original duration was at or below chance level (Fig. 2, dashed line), indicating that the most compressed speech sentences were unintelligible.

The neurosurgical subject patients (Fig. 2, symbols) were not considerably different from a group of tested healthy volunteers (Fig. 2, lines with error bars) in terms of their ability to comprehend time-compressed speech. A two-factor repeated-measures ANOVA was conducted to evaluate the effect of subject popula-

tion and compression ratio on comprehension index. In this repeated-measures design, the between-subject factor was subject population with two levels (patients and volunteers) and within-subject factor was compression ratio with five levels (0.75, 0.50, 0.40, 0.30, 0.20). The α level was set at 0.05. A significant main effect was found for compression ratio ($F_{(4,20)} = 112.42$, $p < 0.0001$). The main effect for subject population was not significant ($F_{(1,23)} = 0.66$, $p < 0.42$), nor were factor interactions ($F_{(4,20)} = 0.072$, $p < 0.58$). The results of this psychophysical test are consistent with speech comprehension data reported previously by Ahissar et al. (2001), obtained using essentially the same experimental paradigm [compare Fig. 3C of Ahissar et al. (2001)].

Cortical responses to time-compressed speech

Time-compressed speech stimuli elicited robust AEPs in HG, with responses having the shortest latencies and highest amplitudes in the posteromedial portion of the gyrus (Fig. 3). Here temporal synchrony to the speech envelope was evident at moderate degrees of compression (0.75 to 0.40) as a series of peaks in the AEP waveform (Fig. 3B, contacts 3–8). At compression ratios that affected comprehension (0.30 to 0.20), however, responses were dominated by a relatively large waveform complex that was time locked to the stimulus onset. Synchrony to the temporal envelope of the stimulus was not apparent. In contrast, AEPs recorded from anterolateral HG (contacts 9–12) had longer latencies, lower amplitudes, and little or no evidence of envelope following.

The AEP waveforms are useful in evaluating the response waveform that is phase locked to the stimulus waveform and

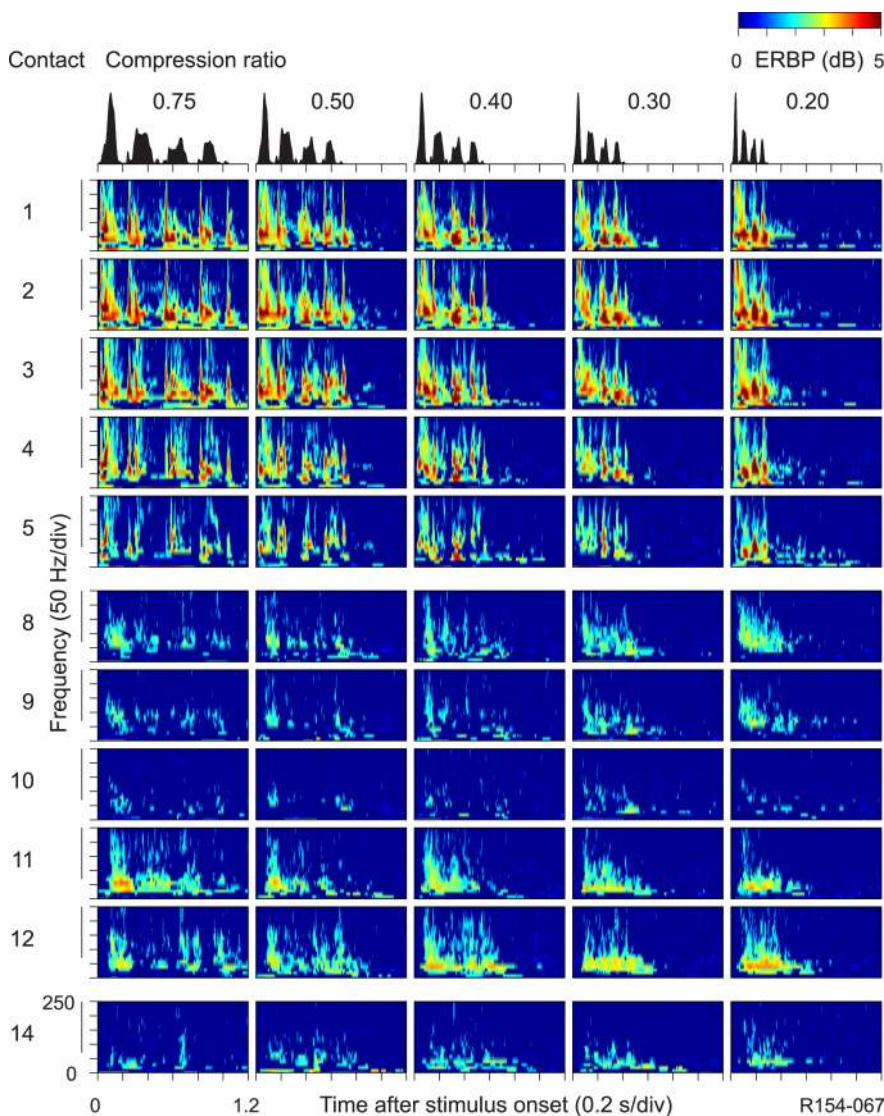


Figure 6. ERBP analysis of recordings from right HG (same subject as in Fig. 5). See legend of Figure 4 for details. Recordings from contacts 6, 7, and 13 were contaminated with power line noise (60 Hz) and are not shown.

largely invariant across trials. Response activity that is time locked but not phase locked to the stimulus waveform would necessarily be markedly attenuated in the across-trial averaging process (Woody, 1967; Glaser and Ruchkin, 1976). To explore this component of speech-evoked activity, we performed spectral analyses of the ECoG data recorded from each of the HG recording sites on a trial-by-trial basis and measured changes in ERBP across a range of frequencies that extended from 10 to 250 Hz (see Materials and Methods). Figure 4 shows the results of such an analysis applied to the dataset introduced in Figure 3. Within posteromedial HG, cortical activity exhibited increases of ERBP that spanned a wide range of frequencies and were most prominent in the high-frequency (70 Hz and above) range. ERBP was not constant in magnitude throughout the duration of the stimulus but appeared to be modulated by the temporal envelope of the speech stimulus. This pattern of ERBP changes, seemingly driven by the stimulus temporal envelope, was observed even in responses to the most compressed (0.30 to 0.20) stimuli at some (Fig. 4, contacts 3–6) but not at all recording sites within HG. This finding is in contrast to the AEP, which failed to

reveal following of the temporal envelope of severely compressed stimuli (compare Fig. 3B).

Additional differences between the ERBP and the AEP include an abrupt decline lateral to contact 6 in the magnitude of the ERBP at all compression ratios, despite the presence of strong temporal synchrony in the AEP at contacts 7 and 8. ERBP data obtained from sites more anterolateral on HG showed some responses throughout the duration of sound stimulus extending to contact 12. These power changes, which were clearly seen between ~50 and 150 Hz, were relatively modest and did not exhibit modulation by the stimulus temporal envelope seen in posteromedial HG, even under the least compressed (0.75) condition.

Representation of abrupt-onset and steady-state components of the speech envelope

ERBP measures shown in Figure 4 show stimulus-related changes in both phase-locked and non-phase-locked power. Abrupt-onset components of complex acoustic stimuli such as speech are likely to trigger cortical activity with a relatively high degree of temporal synchrony. We hypothesized that different components of the speech envelope (such as syllable onsets and vowel nuclei) might be differentially represented by high-frequency cortical activity. To address this question, we attempted to minimize the contribution of phase-locked power by subtracting the AEP from each ECoG trial waveform before time–frequency ERBP analysis (Pulvermüller et al., 1997; Crone et al., 2001). We found that NPL activity in core auditory cortex exhibited modulation by the stimulus envelope (supplemental Fig.

2, available at www.jneurosci.org as supplemental material). On the other hand, syllable onsets were emphasized in the cortical response by phase-locked high-frequency activity, as can be seen from a comparison between plots of total and NPL ERBP in supplemental Figure 2 (available at www.jneurosci.org as supplemental material). Although detailed comparison of phase-locked and NPL high-frequency auditory cortical activity is beyond the scope of this study and currently is under further investigation, we note that phase-locked and NPL ERBP may differentially represent rising and steady-state components of the speech envelope, respectively.

Bilateral responses to the speech envelope

A question remains as to the extent to which temporal synchrony to the speech envelope is represented by posteromedial HG bilaterally (Liégeois-Chauvel et al., 1999, 2004). We could not address this question directly, as simultaneous recording from the left and right hemispheres from the same subjects was not possible due to clinical considerations. We were, however, able to compare data recorded from left (language-

dominant) and right (nondominant) hemispheres across the studied group of subjects.

Results obtained from HG of the right (nondominant) hemisphere in a representative subject are shown in Figures 5 and 6. As with recordings from HG of the language-dominant hemisphere shown previously (Figs. 3, 4), robust AEPs were recorded in posteromedial HG at all compression ratios. Synchrony to the stimulus envelope at compression ratios of 0.75 to 0.50 was most evident at several adjacent recording sites (Fig. 5, contacts 3, 4, 5) located in the central portion of posteromedial HG. There appeared to be a shift in location of the envelope-following response in the AEP (contacts 8, 9, 10) at compression ratios of 0.40 to 0.30. ERBP exhibited modulation by the stimulus envelope at compression ratios extending to 0.20 (Fig. 6). Again, the most prominent temporal modulation of ERBP was in posteromedial HG, and, again, the spatial distribution of ERBP modulation was not entirely coextensive with that of AEP envelope following. A transition seems to have occurred around contact 10, both in the AEP and ERBP. Further anterolaterally on HG (contacts 11 and 12), the AEP was of low magnitude and showed little or no sign of envelope following. ERBP, on the other hand, revealed a faint representation of the stimulus envelope of the least-compressed stimuli (compression ratios 0.75 to 0.50).

Although envelope following was recorded in posteromedial HG in all subjects studied, there was considerable intersubject variability. This is illustrated in Figure 7, which presents AEPs and ERBP envelopes in response to the compressed speech stimuli, recorded at sites of maximal ERBP change within posteromedial HG for all six subjects. In three subjects, recordings were made from right (R), language-nondominant hemisphere, while in three others, data were obtained from left (L), language-dominant hemisphere. In all cases, and at all degrees of compression, stimulus-evoked activity was robust within the high-frequency range (70 Hz and up), peaking at $\sim 3\text{--}6$ dB re prestimulus baseline. Envelope following was also exhibited by all subjects, but the strength of response modulation varied considerably among them even at the lowest degree (0.75) of compression. At the most compressed conditions (0.30 to 0.20), where intelligibility of speech considerably deteriorated (Fig. 2), envelope following was still present in four subjects (L156, R154, R153, and L173).

Representation of the temporal stimulus envelope by the AEP and ERBP

To quantify the representation of the temporal stimulus envelope in the cortical activity, we used two approaches. In the time domain, the accuracy of envelope following was estimated using cross-correlation analysis (Abrams et al., 2008)

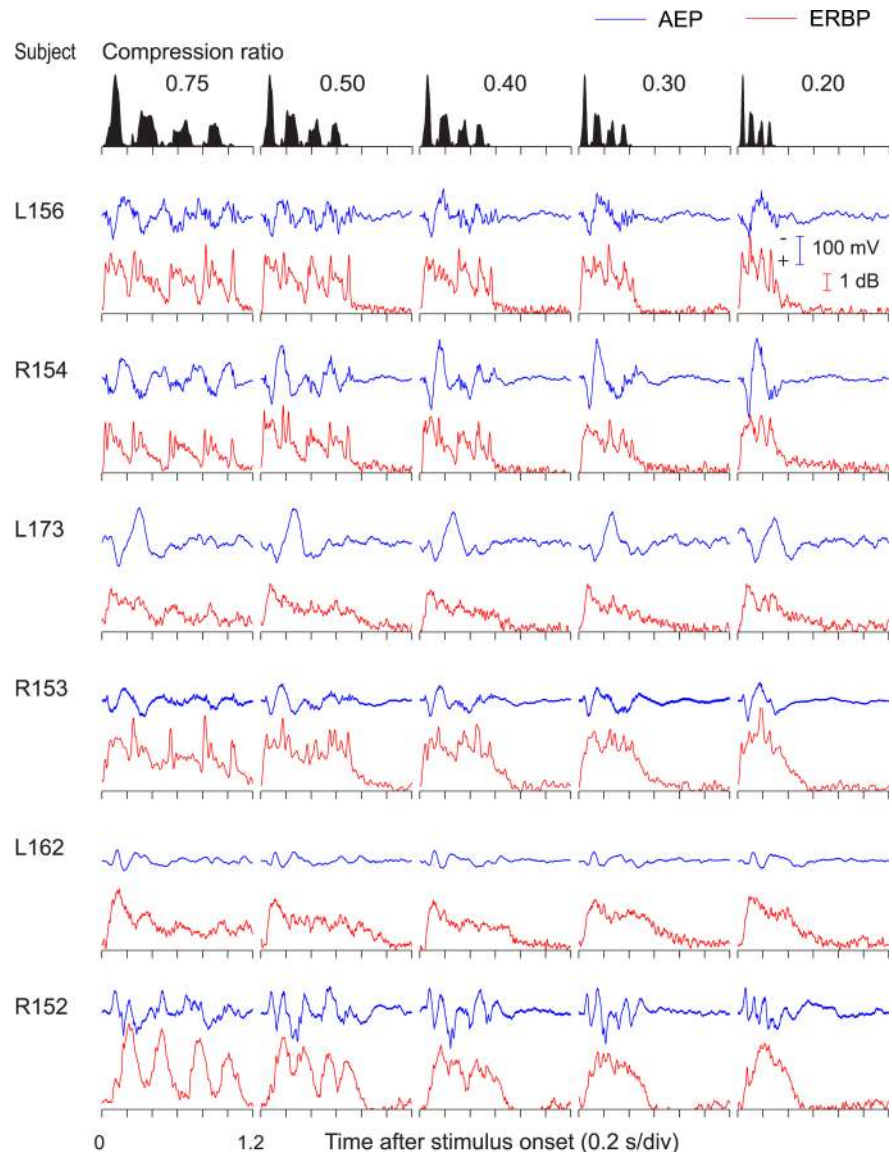


Figure 7. Responses to time-compressed speech sentences recorded from core auditory cortex in six subjects (top to bottom) across compression ratios (left to right: moderate to severe compression). AEPs and ERBP envelopes are plotted in blue and red, respectively. Temporal envelopes of the speech stimuli are shown in the top panels.

and, in the frequency domain, using analysis of stimulus–response modal frequency matching (Ahissar and Ahissar, 2005).

First, envelope following by the AEP and ERBP within core auditory cortex was quantified by measuring peaks of cross-correlograms between speech envelopes and AEPs, and high-frequency ERBP envelope (70–250 Hz; see Materials and Methods), respectively. Figure 8 presents the results of this analysis performed on data obtained from the six subjects at the same core auditory cortex locations as those shown in Figure 7. In four subjects out of six (L156, L173, R154, and R153), correlation between the stimulus envelope and the high-frequency ERBP envelope remained consistently high across compression ratios, including the most compressed (unintelligible) condition. This applies to both total and non-phase-locked ERBP (Fig. 8, open squares and triangles, respectively). In contrast, AEP's stimulus envelope following (Fig. 8, filled circles) deteriorated with compression, consistent with a decrease in comprehension of highly compressed sentences (compare Fig. 2). Similarly, correlation between the stimulus envelope and the ERBP envelope in lower-

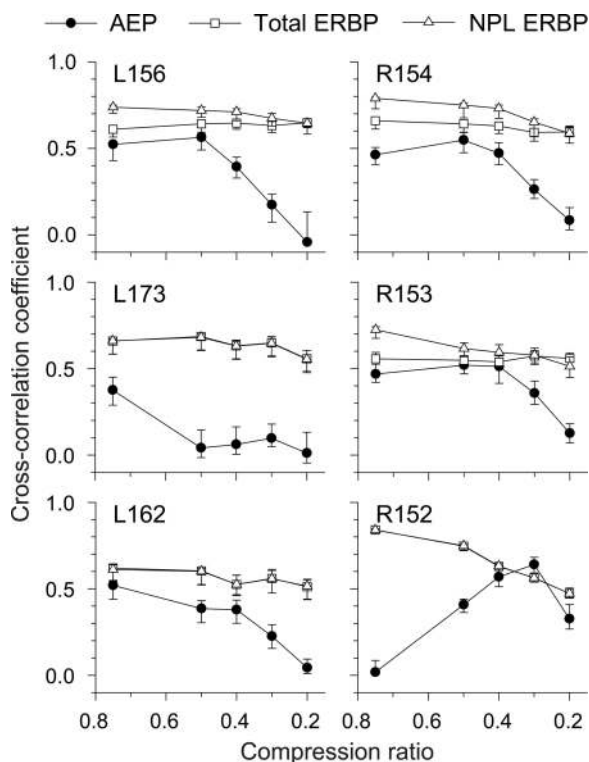


Figure 8. Peak values of cross-correlograms between speech envelopes and AEPs (filled circles), total ERBP envelopes (open squares), and NPL ERBP envelopes (open triangles). Data from six subjects; same contacts as in Figure 5. Error bars indicate 95% confidence intervals.

frequency bands (<50 Hz) did not reliably follow temporal envelope of speech across the range of compression ratios (data not shown).

Next, we sought to examine the extent of frequency matching between the temporal stimulus envelope and recorded cortical activity. As the spectral profiles of the speech envelopes were dominated by modal frequencies ranging from 3.7 to 14 Hz (see supplemental Fig. 1, available at www.jneurosci.org as supplemental material), power spectra of cortical activity were estimated within relatively low (up to 25 Hz)-frequency bands. An example of averaged power spectra of ECoG waveforms recorded from multiple HG sites (Fig. 3A) is shown as blue lines in Figure 9. We also characterized modulation of ERBP by plotting its power spectra across locations and compression ratios (Fig. 9, red lines). The two power measures of the cortical response were compared with the power spectrum of the stimulus envelope (Fig. 9, gray lines).

Power spectra of the ECoG recorded from the posteromedial HG (contacts 3–9) featured peaks that matched the modal frequency of the stimulus envelope at moderate degrees of compression (0.75, 0.50, and 0.40), where speech stimuli were intelligible. This frequency matching was not present, however, in responses to the more compressed stimuli (0.30 and 0.20). In contrast, power spectra of ERBP envelopes exhibited peaks that matched the modal frequency of the stimulus envelopes even in the most compressed condition (0.20) (Fig. 9, contacts 3–6).

Stimulus–response frequency matching was measured in the six subjects as a difference between the modal frequency of the stimulus envelope and a local peak of the response spectrum (Fig. 10). The low-frequency components of the ECoG exhibited frequency matching with the stimulus envelope of sentences compressed to ratios of 0.75, 0.50, and 0.40, and lack of frequency

matching to more compressed stimuli (0.30 to 0.20). This finding is consistent with the MEG data reported by Ahissar et al. (2001). In contrast, the envelope of ERBP exhibited a more accurate frequency matching than low-frequency ECoG components and featured local spectral peaks matching the modal frequency of the stimulus envelope even at compression ratios of 0.30 to 0.20 in four subjects of six (L156, L173, R154, and R153).

We also sought to establish a relationship between comprehension of time-compressed speech and the ability of the core auditory cortex to follow its temporal envelope either by phase-locking of low-frequency ECoG components, or by amplitude modulation of high-frequency activity. For this purpose, we computed correlation coefficients between speech comprehension (measured as comprehension index) (Fig. 2), on the one hand, and the accuracy of cortical envelope tracking (measured in the time domain as peaks of cross-correlations and in the frequency domain as frequency matching) (Figs. 8, 10), on the other. The results are presented in Figure 11. It can be observed that envelope following by low-frequency ECoG components exhibited strong positive correlations with speech comprehension ($r = 0.55$ and $r = 0.66$ for time and frequency domain measures of envelope-following responses). On the other hand, modulation of high-frequency cortical activity produced a generally more faithful representation of the stimulus envelope across speech comprehension, and exhibited only weak positive correlations with it ($r = 0.13$ and $r = 0.19$). This suggests that high-frequency activity within the core auditory cortex can follow the envelope of speech stimuli by and large regardless of their intelligibility.

Discussion

The results of the present study provide evidence that human auditory cortex resolves the temporal envelope of speech stimuli presented at natural speaking rates, as well as at degrees of time compression that make speech unintelligible. It does so by using mechanisms operating over a wide range of ECoG frequencies, at least as high as 250 Hz. This temporal representation is most prominently featured within a restricted region of posteromedial HG, the presumed core auditory cortex, in both dominant and nondominant hemispheres.

Intersubject variability

Despite selecting only those data from electrodes confirmed to be in posteromedial HG gray matter, modulation of both the AEP and the high-frequency ERBP varied among subjects at all rates of utterance. Moreover, in two subjects out of six there was no evidence of envelope following in the ERBP at the most compressed conditions (0.30 to 0.20) (Figs. 8, 10). We note that standard audiometric test results showed speech reception scores in the normal range for all six subjects. Furthermore, all subjects were able to comprehend the speech stimuli when presented at compression ratios between 0.75 and 0.40 (Fig. 2). While cortical high-frequency activity can be influenced by selective attention (Ray et al., 2008), it is unlikely that the attentional load or the degree of arousal contributed to the intersubject variability in cortical response locking in the present study, which used an active-listening task.

Intersubject variability is more likely associated with the functional organization within the core auditory cortex. What we have considered so far to be the core auditory cortex in our human subjects is not expected to be uniform in its cytoarchitecture, nor is it constant relative to gross anatomical landmarks (Galaburda and Sanides, 1980; Rademacher et al., 1993; Leonard et al., 1998;

Hackett, 2003; Fullerton and Pandya, 2007). It is possible that we obtained the data shown from different primary or “primary-like” fields making up the human auditory cortical core. In monkey, three fields within the core have been shown to exhibit demonstrably different capacities to encode temporal information (Bendor and Wang, 2008). Each of the core fields may exhibit tonotopy, which we did not map and which may have influenced temporal synchrony to our speech utterances. For example, the magnitude of the evoked response to a stop consonant is influenced by the onset spectra of the stimulus and where the recording electrode is located within the tonotopic map of the primary auditory cortex (Steinschneider et al., 1995). Other functional organizations may be operating here as well to give rise to intersubject variability (Read et al., 2002).

Comparison with other relevant studies and interpretation of results

At relatively moderate time compression, where the speech utterances were typically intelligible, the AEP showed clear temporal following of the speech envelope. However, when the utterance was accelerated further, envelope following declined, and when the utterance was no longer intelligible, envelope following was no longer in evidence in the AEP. These results are consistent with the findings of Luo and Poeppel (2007), who demonstrated that low-frequency (4–8 Hz) cortical activity measured by MEG was phase locked to the speech signal and that this mechanism correlated with speech intelligibility. Our results are also consistent with the findings of Ahissar et al. (2001) showing that averaged evoked cortical activity measured by MEG was temporally locked to the envelope of moderately compressed speech sentences but failed to synchronize to the envelope of severely accelerated and unintelligible speech. From this they hypothesized that cortical envelope following was a prerequisite for speech comprehension. However, when we analyzed high-frequency ERBP, using the same stimulus paradigm as Ahissar et al. (2001), a more complex picture emerged. Here we observed the ability of the core auditory cortex to synchronize to the speech envelope in the high-frequency range of the ECoG even at rates that made the speech utterance unintelligible. We can therefore conclude that the ability of the auditory core cortex to follow low-frequency fluctuations in the speech envelope is not per se a limiting factor for speech comprehension.

The relationships between response metrics (e.g., AEP, ERBP) derived from ECoG recordings and the intracortical electrodynamicity representing those physiological and behavioral variables assumed to mediate these responses are complex. The literature, spanning more than half a century, is a testament to the importance attached to the resolution of issues that define these relationships (Li and Jasper, 1953; Vaughan and Costa, 1964;

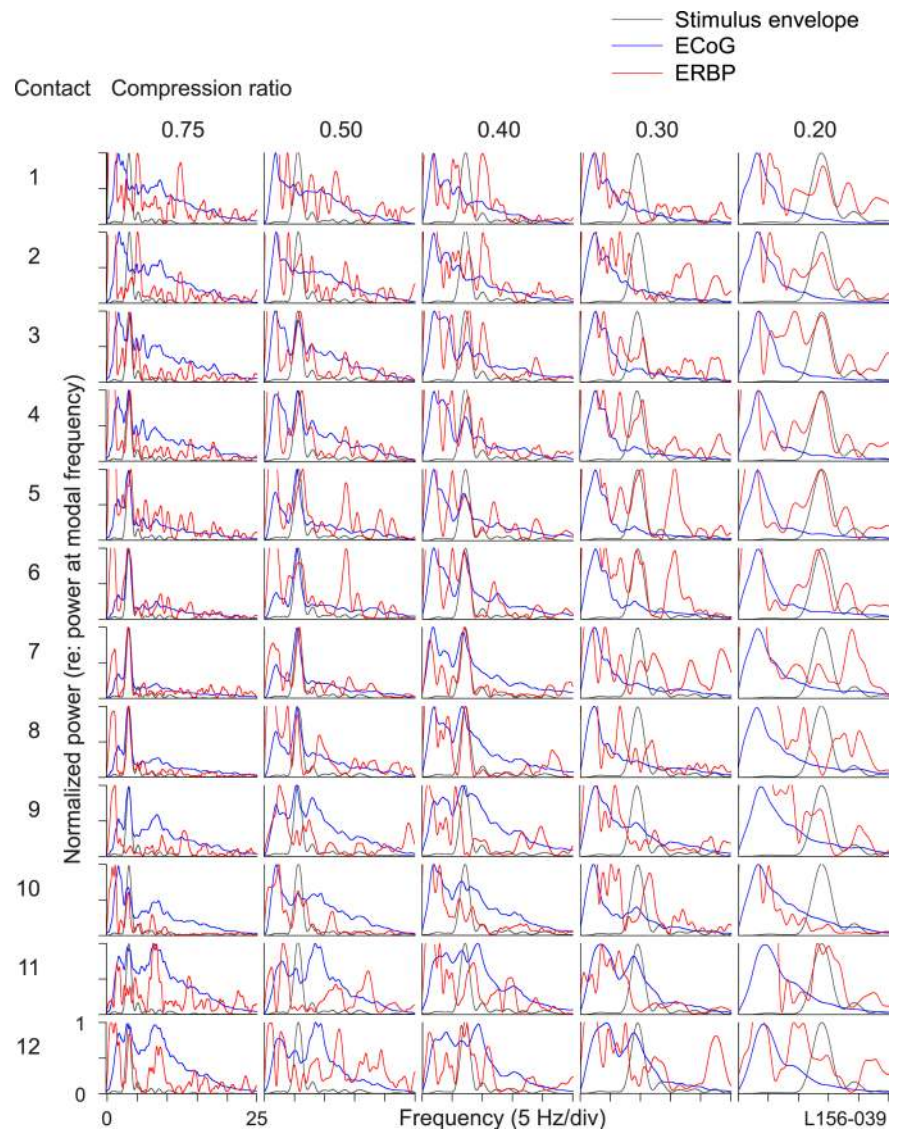


Figure 9. Power spectra of the stimulus envelopes (gray), response waveforms (blue) and ERBP envelopes (red). Data from same subject as in Figure 3 presented across compression ratios (left to right: moderate to severe compression) and the length of HG (top to bottom: posteromedial to anterolateral).

Morrell, 1967; Lopes da Silva et al., 1970; Mitzdorf, 1985; Barth and Di, 1990; Kandel and Buzsáki, 1997; Mukamel et al., 2005; Liu and Newsome, 2006; Ray et al., 2008; Steinschneider et al., 2008; Edwards et al., 2009). Although much progress has been made, there still remains a need for further study to elucidate a definitive and comprehensive explanation. Regardless of whether precise quantitative relationships can be established at this time, we subscribe to the belief that such metrics comprise correlates of time-delimited physiological processes reflecting changes in brain function caused by different stimulus attributes, learning history, and future expectancies whether induced by manifestations of incoming activations, output processes, or memory readouts.

Liégeois-Chauvel et al. (2004) reported greater sensitivity of primary auditory cortex to 4 Hz AM noise in the left hemisphere than in the right hemisphere. Abrams et al. (2008), however, measured scalp electroencephalogram (EEG) responses to slow temporal features of speech corresponding to the syllable rate and found that envelope-following responses were larger on the right hemisphere. Clinical considerations prevented us from recording directly from the left and right hemispheres in the same subject.

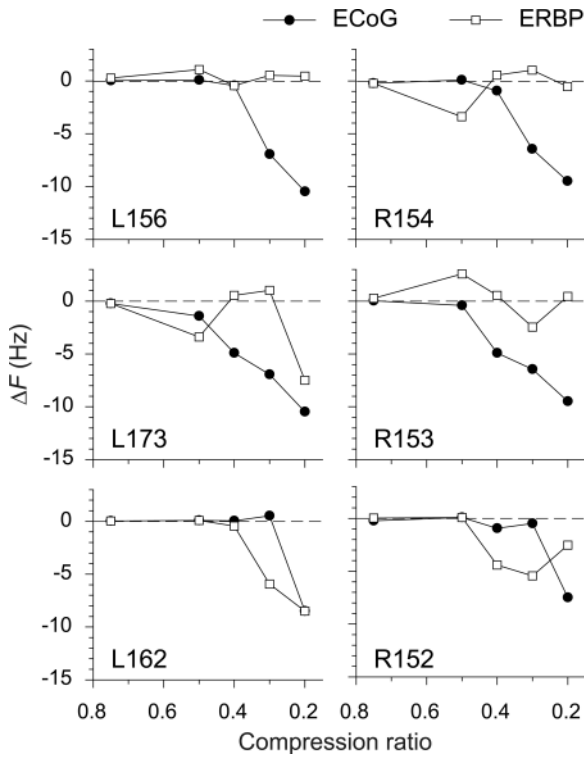


Figure 10. Stimulus–response frequency matching. Filled circles represent frequency difference between the modal frequencies of the stimulus envelope and local maxima of the averaged spectra of ECoG. Open squares represent frequency difference between modal frequencies of the stimulus envelope and local maxima of ERBP spectra. Data from six subjects; same contacts as in Figures 5 and 6.

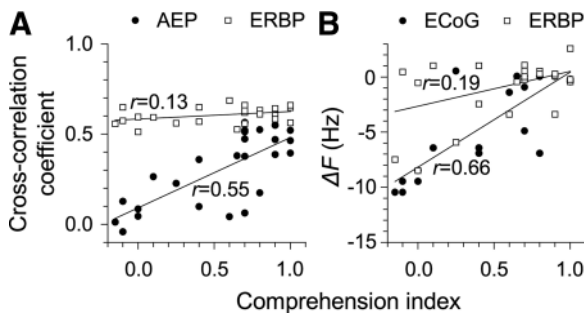


Figure 11. Correlation between envelope following of core auditory cortical responses and speech comprehension. **A**, Peak values of cross-correlograms between speech envelopes and cortical responses (filled circles, AEPs; open squares, ERBP envelopes) are plotted against comprehension index values. **B**, Differences between the modal frequencies of the stimulus envelope and local maxima of the cortical response spectra (filled circles, ECoG; open squares, ERBP envelopes) are plotted against comprehension index values. Lines indicate linear regression based on data from five subjects, in which the psychophysical experiment was performed (L156, L173, L162, R154, and R153).

Nonetheless, taking into account intersubject differences and small sample size, we have no compelling evidence for across-hemisphere differences in the representation of the temporal envelope within HG. Combining functional magnetic resonance imaging and ECoG measures of auditory cortical activity in same subjects may be helpful in addressing this issue.

The speech sentences used by Ahissar et al. (2001) and in the present study can be characterized in terms of the modal frequencies of their temporal envelopes, which correspond to the average syllabic rate. The range of the modal frequencies, from 3.7 to 14

Hz (corresponding to compression ratios of 0.75 and 0.20 of a normal speaking rate), is within the envelope-following capacity of auditory cortex for sinusoidal AM acoustic stimuli, as has been shown previously using fMRI (Giraud et al., 2000) as well as time-domain averaged EEG, MEG, and ECoG recordings (Kuwada et al., 1986; Rees et al., 1986; Roß et al., 2000; Liégeois-Chauvel et al., 2004; Nourski et al., 2009). By compressing the utterance, its duration was reduced from 1.05 to 0.29 s. In the case of the shortest-duration stimulus, a large AEP to stimulus onset may have obscured a phase-locked response. The onset AEP response, however, does not mask high-frequency activity revealed by ERBP measures even with the shortest-duration stimulus.

Periodic acoustic stimuli presented at repetition rates between 3.7 and 14 Hz can elicit percepts of discrete events at the low end and flutter at the high end (Rosen, 1992; Bendor and Wang, 2007). We note that whereas speech comprehension may be lost at high degrees of time compression, the perception of acoustic flutter associated with envelope frequency of the speech signal is indeed retained. The discrepancy between the AEP and the high-frequency ERBP ability to follow the accelerated speech envelope may provide a physiological counterpart of the difference between speech comprehension and perception of its acoustic features at a relatively early level of cortical speech processing within the core auditory cortex.

The current data reveal that the primary auditory cortex is capable of resolving low-frequency (up to at least 15 Hz) envelope information of time-compressed speech, which represents segmental cues. However, cortical evoked activity may not represent information on a shorter time scale (milliseconds to tens of milliseconds) that corresponds to voice onset times and formant transitions and is critical to speech perception. Based on the results of lesion studies, it has been suggested that human primary auditory cortex plays a specialized role in processing speech information in the milliseconds to tens of milliseconds time frame (Phillips and Farmer, 1990). The speech compression method used in the study of Ahissar et al. (2001), and the current study, preserves spectral features of natural speech, but does alter the timing of voice onset and formant transitions. Experiments are underway to examine whether the inability to comprehend compressed speech correlates with a loss of the ability of the primary auditory cortex to temporally represent stimulus features such as voice onsets and formant transitions, occurring on a time scale of tens of milliseconds.

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, Ahissar M (2005) Processing of the temporal envelope of speech. In: *The auditory cortex: a synthesis of human and animal research* (König R, Heil P, Budinger E, Scheich H, eds), pp 295–314. Mahwah, NJ: Erlbaum.
 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.
 Barth DS, Di S (1990) Three-dimensional analysis of auditory-evoked potentials in rat neocortex. *J Neurophysiol* 64:1527–1536.
 Bendor D, Wang X (2007) Differential neural coding of acoustic flutter within primate auditory cortex. *Nat Neurosci* 10:763–771.
 Bendor D, Wang X (2008) Neural response properties of primary, rostral, and rostromedial core fields in the auditory cortex of marmoset monkeys. *J Neurophysiol* 100:888–906.
 Bieser A, Müller-Preuss P (1996) Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. *Exp Brain Res* 108:273–284.

- Brugge JF, Volkov IO, Oya H, Kawasaki H, Reale RA, Fenoy A, Steinschneider M, Howard MA 3rd (2008) Functional localization of auditory cortical fields of human: click-train stimulation. *Hear Res* 238:12–24.
- Crone NE, Miglioretti DL, Gordon B, Lesser RP (1998) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* 121:2301–2315.
- Crone NE, Boatman D, Gordon B, Hao L (2001) Induced electrocorticographic gamma activity during auditory perception. *Clin Neurophysiol* 112:565–582.
- Donchin DB, Lindsley E (1969) Average evoked potentials: methods, results and evaluations. Washington, DC: NASA.
- Drullman R, Festen JM, Plomp R (1994) Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am* 95:1053–1064.
- Edwards E, Soltani M, Kim W, Dalal SS, Nagarajan SS, Berger MS, Knight RT (2009) Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. *J Neurophysiol* 102:377–386.
- Fullerton BC, Pandya DN (2007) Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. *J Comp Neurol* 504:470–498.
- Galaburda A, Sanides F (1980) Cytoarchitectonic organization of the human auditory cortex. *J Comp Neurol* 190:597–610.
- Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A (2000) Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol* 84:1588–1598.
- Glaser EM, Ruchkin DS (1976) Principles of neurobiological signal analysis. New York: Academic.
- Hackett TA (2003) The comparative anatomy of the primate auditory cortex. In: *Primate audition: ethology and neurobiology* (Ghazanfar AA, ed), pp 199–219. Boca Raton, FL: CRC.
- Howard MA 3rd, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA (1996) A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res* 724:260–264.
- Howard MA, Volkov IO, Mirsky R, Garell PC, Noh MD, Granner M, Damasio H, Steinschneider M, Reale RA, Hind JE, Brugge JF (2000) Auditory cortex on the human posterior superior temporal gyrus. *J Comp Neurol* 416:79–92.
- Kalcher J, Pfurtscheller G (1995) Discrimination between phase-locked and non-phase-locked event-related EEG activity. *Electroencephalogr Clin Neurophysiol* 94:381–384.
- Kandel A, Buzsáki G (1997) Cellular-synaptic generation of sleep spindles, spike-and-wave discharges, and evoked thalamocortical responses in the neocortex of the rat. *J Neurosci* 17:6783–6797.
- Knuth KH, Shah AS, Truccolo WA, Ding M, Bressler SL, Schroeder CE (2006) Differentially variable component analysis: identifying multiple evoked components using trial-to-trial variability. *J Neurophysiol* 95:3257–3276.
- Kuwada S, Batra R, Maher VL (1986) Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones. *Hear Res* 21:179–192.
- Leonard CM, Puranik C, Kuldau JM, Lombardino LJ (1998) Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? *Cereb Cortex* 8:397–406.
- Li CL, Jasper H (1953) Microelectrode studies of the electrical activity of the cerebral cortex in the cat. *J Physiol* 121:117–140.
- Liegeois-Chauvel C, Musolino A, Chauvel P (1991) Localization of the primary auditory area in man. *Brain* 114:139–151.
- Liégeois-Chauvel C, de Graaf JB, Laguitton V, Chauvel P (1999) Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cereb Cortex* 9:484–496.
- Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J, Chauvel P (2004) Temporal envelope processing in the left and right auditory cortices. *Cereb Cortex* 14:731–740.
- Liu J, Newsome WT (2006) Local field potential in cortical area MT: stimulus tuning and behavioral correlations. *J Neurosci* 26:7779–7790.
- Lopes da Silva FH, van Rotterdam A, Storm van Leeuwen W, Tielens AM (1970) Dynamic characteristics of visual evoked potentials in the dog. I. Cortical and subcortical potentials evoked by sine wave modulated light. *Electroencephalogr Clin Neurophysiol* 29:246–259.
- Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001–1010.
- Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol Rev* 65:37–100.
- Morrell LK (1967) Temporal characteristics of sensory interaction: evoked potential and reaction time observations. *Electroencephalogr Clin Neurophysiol* 23:77.
- Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R (2005) Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309:951–954.
- Nourski K, Oya H, Kawasaki H, Reale R, Chen H, Howard M, Brugge J (2009) Representation of temporal sound features by high frequency gamma activity in the human core auditory cortex. *Assoc Res Otolaryngol Abstr* 298.
- Oya H, Kawasaki H, Howard MA 3rd, Adolphs R (2002) Electrophysiological responses in the human amygdala discriminate emotion categories of complex visual stimuli. *J Neurosci* 22:9502–9512.
- Pantev C (1995) Evoked and induced gamma-band activity of the human cortex. *Brain Topogr* 7:321–330.
- Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin Neurophysiol* 110:1842–1857.
- Phillips DP, Farmer ME (1990) Acquired word deafness, and the temporal grain of sound representation in the primary auditory cortex. *Behav Brain Res* 40:85–94.
- Pulvermüller F, Birbaumer N, Luttenberger W, Mohr B (1997) High-frequency brain activity: its possible role in attention, perception and language processing. *Prog Neurobiol* 52:427–445.
- Rademacher J, Caviness VS Jr, Steinmetz H, Galaburda AM (1993) Topographical variation of the human primary cortices; implications for neuroimaging, brain mapping and neurobiology. *Cereb Cortex* 3:313–329.
- Ray S, Niebur E, Hsiao SS, Sinai A, Crone NE (2008) High-frequency gamma activity (80–150 Hz) is increased in human cortex during selective attention. *Clin Neurophysiol* 119:116–133.
- Read HL, Winer JA, Schreiner CE (2002) Functional architecture of auditory cortex. *Curr Opin Neurobiol* 12:433–440.
- Reddy CG, Dahdaleh NS, Albert G, Chen F, Hansen D, Nourski K, Kawasaki H, Oya H, Howard MA 3rd (2009) A method for placing Heschl gyrus depth electrodes. Technical note. *J Neurosurg*. Advance online publication. Retrieved Nov. 24, 2009. doi:10.3171/2009.7.JNS09404.
- Rees A, Green GG, Kay RH (1986) Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear Res* 23:123–133.
- Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci* 336:367–373.
- Roß B, Borgmann C, Draganova R, Roberts LE, Pantev C (2000) A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. *J Acoust Soc Am* 108:679–691.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270:303–304.
- Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416:87–90.
- Steinschneider M, Reser D, Schroeder CE, Arezzo JC (1995) Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain Res* 674:147–152.
- Steinschneider M, Fishman YI, Arezzo JC (2008) Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cereb Cortex* 18:610–625.
- Thomson DJ (1982) Spectrum estimation and harmonic analysis. *Proc IEEE* 70:1055–1096.
- Vaughan HG Jr, Costa LD (1964) Application of evoked potential techniques to behavioral investigation. *Ann NY Acad Sci* 118:71–75.
- Woody CD (1967) Characterization of an adaptive filter for the characterization of variable latency neuroelectric signals. *Med Biol Eng* 5:539–553.