Comparison of Time–Frequency Responses and the Event-Related Potential to Auditory Speech Stimuli in Human Cortex

Erik Edwards,1,* Maryam Soltani,1,* Won Kim,1 Sarang S. Dalal,2 Srikantan S. Nagarajan,2 Mitchel S. Berger,3 and Robert T. Knight1

1Department of Psychology and Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley; and 2Department of Radiology and Biomedical Imaging and 3Department of Neurological Surgery, University of California, San Francisco, California

Submitted 23 August 2008; accepted in final form 10 May 2009

Edwards E, Soltani M, Kim W, Dalal SS, Nagarajan SS, Berger MS, Knight RT. Comparison of time–frequency responses and the event-related potential to auditory speech stimuli in human cortex. J Neurophysiol 102: 377–386, 2009. First published May 13, 2009; doi:10.1152/jn.90954.2008. We recorded the electrocorticogram directly from the exposed cortical surface of awake neurosurgical patients during the presentation of auditory syllable stimuli. All patients were unanesthetized as part of a language-mapping procedure for subsequent left-hemisphere tumor resection. Time–frequency analyses showed significant high-gamma (γhigh: 70–160 Hz) responses from the left superior temporal gyrus, but no reliable response from the left inferior frontal gyrus. Alpha suppression (α: 7–14 Hz) and event-related potential responses exhibited a more widespread topography. Across electrodes, the α suppression from 200 to 450 ms correlated with the preceding (50–200 ms) γhigh increase. The results are discussed in terms of the different physiological origins of these electrocortical signals.

INTRODUCTION

Long before microelectrode recordings of spiking in single cortical neurons, electrophysiological exploration of cortex used electrodes placed on the cortical surface of animals (Beck 1891; Caton 1875), a technique now known as electrocorticography (ECoG). Although the early pioneers of animal ECoG were primarily concerned with evoked changes in potential level (Beck 1891; Caton 1875), they had also noted the existence of spontaneous fluctuations (Brazier 1961). Beck (1891) was the first to report suppression of spontaneous ECoG fluctuations to a simple sensory stimulus, a finding now termed “event-related desynchronization” (Pfurtscheller and Aranibar 1977). This finding was confirmed in dogs by Pravdich-Nemisky (1925) and in the human scalp electroencephalogram (EEG) by Berger (1930) who noted the suppression to be centered in the alpha band (α, 7–15 Hz by current convention). The outstanding early study of the animal ECoG was by Ectors (1936) who demonstrated the three major aspects of the typical ECoG response to activating stimuli and their utility for cortical mapping by systematically studying them in several sensory and motor areas. The three aspects of this ECoG response are 1) an evoked potential, now typically studied with event-related averaging and called the event-related potential (ERP); 2) a suppression of the spontaneous low-frequency (<30 Hz) activity; and 3) an increase in today’s “gamma” range (γ, >30 Hz). Ectors also noted frequencies as high as 80–100 Hz for stimuli that were particularly strong or effective and was thus the first to describe today’s “high-gamma” band (γhigh, studied here from 70 to 160 Hz). The results of Ectors (1936) and these three major categories of ECoG response to activating events have been abundantly confirmed in subsequent animal and human studies, including recent studies using time–frequency (TF) analyses not available in that era. In particular, TF analyses of human ECoG data have recently indicated the γhigh band (~60–300 Hz, studied here from 70 to 160 Hz) as a powerful means of mapping task-specific cortical activation (Brovelli et al. 2005; Canolty et al. 2007; Crone et al. 1998, 2001; Edwards et al. 2005).

There were two overall goals of this study. The first goal was to more carefully study the relationships between the different electrophysiological measures—the ERP and the TF responses in different frequency bands—and we use the intertrial coherence (ITC) to study the phase-consistent TF responses that contribute to the ERP. Using the unique advantages of intracranial recordings from human cortex, we asked to what extent these different measures are correlated and localized. The physiological origins of these different measures remain poorly understood, despite their long history, and a study of their interrelations in a simple sensory task may provide insight in this regard (Steinschneider et al. 2008).

The second goal was to identify the location of cortical areas responding to simple speech stimuli in different phonological categories and to characterize the electrophysiological responses to such stimuli in lateral auditory cortices. Individual speech sounds—phonemes—are categorically perceived (Liberman et al. 1957) and it is generally believed that categorical perception and other aspects of auditory speech processing occur at the cortical level (Soltani 2007). The mismatch-negativity (MMN) and its magnetic equivalent have been used successfully in scalp EEG and magnetoencephalographic (MEG) studies of categorical phoneme perception (Aaltosen et al. 1987; Dehaene-Lambertz 1997; Shiestakova et al. 2002). We used the MMN paradigm here, as we did in our prior ECoG study of the tone MMN (Edwards et al. 2005), to examine responses to oddball auditory stimuli that differ in phonetic category but included allophonic and speaker variability. Although negative results are obtained regarding straightforward topographic representation of categorical perception, we are still able to characterize the overall response and topography to simple speech stimuli in higher-order auditory cortical areas.
METHODS

Neurosurgical patients

All patients were undergoing awake language mapping during neurosurgery for brain tumor resection (for more details about the neurosurgical procedure see Berger 1996; Maciunas et al. 1996). They consented to participate in this 10- to 15-min study after language mapping. The experiments were approved by the Committees for the Protection of Human Subjects at UCSF and UC Berkeley and posed no additional risk. We recorded from nine patients (five females, four males, age range 19–57 yr). All patients were right-handed and assumed to be left-hemisphere dominant for language. All craniotomies were over the left hemisphere and all were first time surgeries. Patients were initially anesthetized with a propofol remifentanil combination, but all medication was discontinued before the language mapping prior to our study. There were no known auditory or language deficits in any of the patients. The right ear facing downward was not occluded, whereas the left ear was covered by surgical draping. The patients’ hearing was not significantly compromised, allowing normal conversation with the patients during the language mapping.

Auditory stimuli and experimental design

Five syllables were used in this experiment: /pa/, /ba/, /wa/, /aa/, and /uu/. Syllables were taken from recordings of four male speakers. Each syllable was repeated 20 times by each speaker and, of these, 5 were selected per speaker and per syllable (giving 5 × 5 × 4 = 100 unique sounds). The selection attempted to minimize duration and loudness outliers. Loudness variability was further reduced by scaling the amplitude of each recorded syllable such that its root-mean-square (rms) level was brought one third closer to the mean rms over all syllables. This reduced the SD of rms levels by 33%, but we did not want identical rms levels because our experimental design called for some acoustical variability.

The syllables were presented to subjects in two blocks of 500 stimuli, each block lasting about 6.5 min. The order of blocks was counterbalanced across subjects. In two patients, recording was stopped after the first block, which was the MMN block in both patients. Background noise level in the operating room was about 50–55 dB SPL; beeps of the heart rate monitor were turned off for the language-mapping procedure and the experimental session. Syllables were presented at about 70–75 dB SPL via speakers placed in front and below the patient’s head at about 50-cm distance. The patients were instructed to ignore the syllables and look at a slide presentation of photographs.

In the control block, all five syllables—/pa/, /ba/, /wa/, /aa/, and /uu/—were each presented with 20% probability in pseudorandom order. In the MMN block, the stimulus sequence was identical except that all instances of /wa/, /ba/, and /aa/ were changed to /aa/. Thus /aa/ served as the “standard” (80%) and /aa/ as the “deviant” (20%). The stimulus onset asynchrony (SOA, onset-to-onset) was 725–825 ms. A

Data analyses

All data were imported into MATLAB (The MathWorks, Natick, MA) for processing. The first three stimuli of each block were rejected to allow auditory responses to stabilize. Artifact rejection began with an automatic procedure to detect amplifier saturation or excessive power across all frequency bands indicating a transient artifact. These were checked by visual inspection and additional rejections added manually using EEGLAB (Delorme and Makeig 2004). Less than 5% of trials were rejected by these criteria. After artifact rejections, there were between N = 977 and N = 993 stimuli included in each average (between N = 326 and N = 488 stimuli included for the two patients with only one block).

FIG. 1. Illustration of experimental design for auditory stimulus presentation. In all, 500 stimuli (duration 200–284 ms) were presented in each of 2 blocks with a stimulus onset asynchrony (SOA) of 725–825 ms. A: in the control block, each syllable (/uul, /uul, /wu/, /ba/, /pa/) was presented with equal frequency (20%) in pseudorandom order. B: in the mismatch-negativity (MMN) block, the /aa/ stimuli served as the standard (80%) and the /uu/ stimuli served as the deviant (20%). The sequence was made by replacing all /wa/, /ba/, and /aa/ stimuli from the control block with an /aa/ stimulus. Thus the /uu/ stimuli were identical in the 2 blocks, including their timing in the stimulus sequence.
Event-related potentials (ERPs) were made by taking the average across trials time-locked to stimulus onset. Epochs were taken from 150 ms before stimulus onset to 625 ms after (150 + 625 = 775 ms, the median SOA). Because some channels exhibited a pseudo steady-state response—that is, slower shifts in potential that followed the SOA periodicity (~1.3 Hz) and did not return to baseline before the next stimulus—all channels were high-pass filtered at 2.3 Hz prior to making the ERP. This used a symmetrical, finite-impulse response filter (~35 dB/octave roll-off; >10 dB reduction <2 Hz) from EEGLAB (Delorme and Makeig 2004). This has the additional advantage of eliminating any cardiac pulse artifacts that may be present (60–120 beats per min = 1–2 Hz). The middle- and long-latency auditory evoked potentials of interest, as well as the MMN, are composed of frequencies >3 Hz and are therefore unaffected by this filtering.

TF analyses used a Gaussian filter bank and the Hilbert transform on each channel separately. There is one Gaussian filter (Gaussian-shaped window in the frequency domain) in the filter bank for each of 42 center frequencies (cf) from 4 to 250 Hz. For each cf, the result is a time series of complex numbers called the analytic signal (AS). The absolute value of the AS is the analytic amplitude (AA), which is the envelope of the band-pass–filtered signal. This method is formally equivalent to complex demodulation (Flanagan 1980; Ktonas and Papp 1980; Walter 1968) and to deconvolution with Gaussian-tapered (“Morlet”) wavelets (Bruns 2004), with the caveat that bandwidth (σf) parameters must be matched to obtain identical results. For wavelets, the parameters are typically set so that there is a constant ratio of σf to cf (“constant-Q wavelets”). Thus a plot of log σf versus log cf gives a straight line with slope = 1. For the short-term fast Fourier transform (ST-FFT) approach (Edwards et al. 2005; Pfurtscheller and Aranibar 1977), the same σf is used at all cf, giving slope = 0 for log σf versus log cf. The filter-bank parameters used here fall directly in between, using slope = 0.5 for log σf versus log cf (for Gaussian filters as used here, the SD parameter directly gives the σf parameter of the uncertainty relation). The TF settings were initially approximated by a data-driven, “efficient coding” approach (Lewicki 2002) to identify optimal basis functions. However, some degree of trial and error is unavoidable, even after the use of a data-driven optimization procedure, given that we want to emphasize temporal over frequency resolution for the present task using fairly rapid stimulus presentation. The exact settings were σf = 0.78 Hz at cf = 4 Hz and increasing linearly (on a log–log scale) with slope = 0.5 to 6.17 Hz at cf = 250 Hz.

The resulting AA time series has the same sampling rate (2,003 Hz), duration (~6.5 min), and units (microvolt) as those of the original ECoG recording. Thus event-related and statistical analyses are done identically as described earlier for the ERPs. The AA results are shown in units of percentage (%) changes from the baseline mean. The reason for converting to % units is so that all frequency bands can be displayed together in one color scale despite large differences in their absolute microvolt levels. The conversion to % units is purely a linear and does not distort the AA time series, so it amounts to only a relabeling of the scale.

**Intertrial coherence**

The calculation of intertrial coherence (ITC) (Delorme and Makeig 2004; Edwards et al. 2005; Makeig et al. 2002) uses the complex numbers of the AS. Each complex number is a vector in the complex plane, where the length is the AA and the angle with respect to the positive real axis is the phase. ITC is a measure of the phase alignment of these vectors across trials for a given latency. ITC is a unitless value ranging from 0 (random phase alignment) to 1 (perfect phase alignment). To calculate ITC, the AS vectors from each single trial are normalized to unit length and then averaged across trials (see Eq. 8 of Bruns 2004). The length of the resulting vector is the ITC. If there is a high degree of phase consistency across trials, the unit vectors will align and average to a larger vector than if they are scattered around the phase circle. The ITC is identical to other “phase-locking” measures (Bruns 2004; Lachaux et al. 1999) and high versus low ITC corresponds to “evoked” versus “induced” activity.

The ITC must be corrected for the number of trials N because a nonzero coherence is expected by chance and this chance coherence is larger for smaller N (Benignus 1969; Bruns 2004). To see this, consider N unit vectors scattered randomly around the phase circle, i.e., consider that the phases are drawn randomly from a uniform distribution from 0 to 2π. It is entirely improbable that the vectors will be perfectly arranged around the circle so as to sum to 0, nor perfectly aligned to sum to 1. Rather the coherence expected in this completely random situation is approximately 1/\sqrt{N}. More precisely, chance coherence is distributed according to a Rayleigh with parameter β = 1/\sqrt{N} (Edwards 2007). To correct for the effect of N in coherence, one can subtract the bias-correction factor suggested by Benignus (1969). For a given electrode, we denote the ITC as a function of center frequency (cf) and time (t) and the bias correction, operating on squared coherence, is

\[
\text{ITC}^2(cf, t) = \text{ITC}^2(cf, t) - \frac{1 - \text{ITC}^2(cf, t)}{N}
\]

Very low coherence values are sometimes bias-corrected to negative values—an impossible result—and so these are set to 0. The square root is then taken to give ITC(cf, t), which is used here for all results. Benignus (1969) noted that the bias-correction factor was not entirely accurate for very low coherence values and we will be using the very low ITC values for νhigh range in correlation analysis with the ERP rms levels. Since both the ITC and the ERP rms are biased toward larger values for smaller N, there will be an artificial correlation introduced by differencing N across electrodes. Because the ITC values are very small for the νhigh range and because N varies over a fairly wide range (between 326 and 993 stimuli), the correlation between νhigh ITC and ERP rms must be computed as a partial correlation controlling for 1/\sqrt{N}. That is, the νhigh ITC at any given electrode is essentially at chance (1/\sqrt{N}) and so a correlation with νhigh ITC will in effect be a correlation with 1/\sqrt{N}. It is therefore the partial correlation that we report in the following text.

**Statistical assessment of ERPs**

To assess whether the poststimulus ERP differed significantly from the prestimulus baseline (the within-condition question), we used the resampling procedure of our prior study (Edwards et al. 2005), based on the method of Makeig and colleagues (Delorme and Makeig 2004; Makeig et al. 2002). For each of 2,000 iterations, data points are drawn at random from the prestimulus baseline periods of the single-trial data and averaged together. Each resampled average uses the same number of data points as included in the original ERP. The distribution of these 2,000 averages forms the null distribution and raw P values are obtained as percentiles within this null distribution. Raw P values are corrected for multiple comparisons using the false-discovery rate (FDR) approach (Benjamini and Hochberg 1995; Edwards et al. 2005; Nichols and Hayasaka 2003). In all statistical analyses, the ERPs are first reduced to a sampling rate of 250 Hz to reduce the number of hypotheses tested.

The comparison of ERPs across conditions (i.e., standards vs. deviants) also used a bootstrap resampling procedure. Because no significant differences were obtained for comparisons across conditions (see following text), further details are omitted.

**Statistical assessment of TF analyses**

There is no generally agreed on method for statistical assessment in the TF plane. For each subject, there are 4–7 electrodes (spatial samples), 42 frequency bands in the filter bank, and 1,553 time points in the
peristimulus epoch, resulting in a total of 260,904 to 456,582 space–time–frequency “voxels” to potentially test for significance. Our approach, like most approaches to this problem in EEG/MEG and functional magnetic resonance imaging (fMRI) statistics (Frackowiak 2004), is a “mass univariate” approach. That is, each voxel is assessed independently by some univariate measure (a multivariate measure would assess voxel clusters together) and the entire “mass” of raw \( P \) values is then submitted to a procedure of correction for multiple comparisons.

One problem with this approach is that the total number of voxels tested influences the \( P \) value at every individual voxel by multiple comparison correction; however, the total number of voxels depends on the parameters chosen for the TF analysis. For example, what if more or fewer frequency bands had been included in the filter bank? or what if the sampling rate happened to be different? Although the complete answers to these questions are not agreed on, we introduce one straightforward improvement here: it is important not to oversample in the time or frequency dimensions, leading to more time–frequency voxels than are needed to represent the same information.

In the frequency domain, it can be shown that a Gaussian filter should be separated by at least one \( \sigma_f \) from the neighboring Gaussian filters (Edwards 2007; Singh and Theunissen 2003). Any closer spacing would not lead to any new information and would thus constitute oversampling in the frequency domain. The Gaussian filter bank used here conforms to this theoretic lower limit of \( \sigma_f \) spacing and therefore controls the number of frequency bands submitted to hypothesis testing.

In the time domain, we must consider that the amplitude modulations of a band-limited signal cannot occur at arbitrarily fast frequencies. Indeed, a major motivation behind the mathematical construction of the AS is to place the low frequencies of the signal in the “envelope” (AA) and the high frequencies in the “carrier” (real part of the AS) (Cohen 1995; Flanagan 1980). Thus if the center frequency is “envelope” (AA) and the high frequencies in the “carrier” (real part of the AS) is to place the low frequencies of the signal in the


correction controlling for 1/\( \sqrt{N} \). In the present study, this really affected only correlations involving the ITC, with almost no effect of \( N \) found for correlations involving AA (\( \alpha \), \( \gamma_{\text{high}} \)). Therefore for simplicity we report the straight correlations in the following text except for those involving the ITC, where the partial correlation is used.

**RESULTS**

**Responses from mid-STG**

The overall responses to syllable stimuli are given without distinguishing phoneme categories or standards versus deviants. TF and ERP averages here include all stimuli (range: 326 to 993 stimuli, depending on number of blocks run and artifact rejections), with the aim of characterizing the general pattern of reactivity across frequency bands and cortical areas. Al-
though many electrodes exhibited suppression of lower frequencies (<30 Hz) and ERPs of various polarity and morphology, only electrodes on the mid-STG exhibited robust $\gamma_{\text{high}}$ (70–160 Hz) responses. Note, however, that we lack coverage of the posterior STG and superior temporal plane (STP) that are also expected to give strong responses.

The three most strongly responding electrodes are shown in Fig. 2. All three electrodes exhibit the typical TF response (Fig. 2B) for specifically activated cortical areas—there is suppression of $\alpha$ and lower frequencies $\leq$30 Hz and an increase of higher frequencies $\geq$30 Hz. The border between frequency bands exhibiting decreases versus increases ranged from 25 to 40 Hz, i.e., within the $\beta$ (15–30 Hz) and $\gamma_{\text{low}}$ (30–60 Hz) bands. The peak increase in $\gamma_{\text{high}}$ AA relative to prestimulus baseline occurred at an average latency of 161 ms and at an average frequency of maximum response of 104 Hz (mean for these three electrodes; see legend to Fig. 2 for each individual electrode).

Another typical feature of the TF response at activated sites is an initial increase in amplitude from about 0 to 150 ms, particularly in the frequencies 4–7 Hz (Patients 1 and 6) or 4–15 Hz (Patient 8). This usually corresponds to ERPs occurring during the same latency range, although ERPs will also appear if there is a strong ITC during the same latency range even without an amplitude increase. Figure 2C shows the ITC results for the three mid-STG electrodes and Fig. 2D shows the ERPs. The results for the ERP and the ITC are given in their own section in the following text, but here we note that the low-frequency (4–15 Hz) activity in AA is robustly correlated with ERP rms across electrodes ($r = 0.50$, $P < 0.01$).

**Topography of $\gamma_{\text{high}}$ and $\alpha$ responses**

Results for all nine patients (45 electrodes) are summarized in Fig. 3. Only STG electrodes exhibited significant ($P < 0.05$) $\gamma_{\text{high}}$ AA responses exceeding baseline levels by $>5\%$ (Fig. 3A). The remaining anatomical regions exhibited no or only very weak ($<5\%$ increase) $\gamma_{\text{high}}$ responses, although some of these were significant at the $P < 0.05$ level (see Table 1). By contrast, $\alpha$ suppressions (averaged here from 200 to 450 ms,
where the trough of α suppressions occurred) were observed with a more widespread topography including STG, middle temporal gyrus (MTG), temporoparietal junction (TPJ), and inferior frontal gyrus (IFG) electrodes (Fig. 3B). Over half of the electrodes (30 of 45) exhibited significant α suppression. Thus the topography of α suppression was similar to γhigh increases in that its focus was over the mid-STG, but was dissimilar in showing strong (less than −10%) and highly significant (P < 0.01) responses at more widespread locations. As noted earlier, however, weak and occasionally significant γhigh increases were observed at more widespread sites and we sought to quantify a possible correlation between these two measures.

We assessed the topographical correspondence of α suppression and γhigh increases by a simple correlation analysis across all 45 electrodes (Fig. 3E). By confirming the visual impression from the topography plots, there was a significant negative correlation (r = −0.67, P < 0.01) between average γhigh AA in the latency range 50–200 ms and average α AA in the latency range 200–450 ms. The exact latency ranges chosen (50–200 ms for γhigh and 200–450 ms for α) do not substantially affect the result and these ranges were chosen to emphasize that the correlation does not depend on overlap of the latency ranges. In fact, the correlation decreases slightly by overlapping the two latency ranges, emphasizing that the correlation is primarily between an earlier γhigh increase and a later α suppression. This correlation is not driven by the three “outlier” electrodes from Fig. 2, as indicated by removing these from the analysis. By using only the remaining 42 electrodes, the correlation is only slightly affected and remains significant (r = −0.67, P < 0.01). This is remarkable given that the γhigh increases for these electrodes never exceed 5% and are usually nonsignificant (P > 0.05). This indicates that weak γhigh increases accompany (and precede) α suppressions, even at electrodes where no significant γhigh responses are otherwise noticed.

**ERPs and corresponding TF measures**

ERPs of various polarity and morphology were obtained with a focus of energy in the latency range 50–200 ms (see examples from the mid-STG in Fig. 2D). ERPs exhibited a widespread topography covering STG and MTG sites, but the topography was not identical to α suppression and did not include any IFG sites (Fig. 3C). There was nonetheless a significant correlation between ERPs and α levels across electrodes (r = −0.47, P < 0.01). ERP topography and correlation results are similar using peak levels instead of rms levels.

In contrast to the topographical correspondence between γhigh and α responses, γhigh and ERP responses were not strongly linked. Although there is a positive correlation between γhigh AA and ERP rms levels (r = 0.48), it does not reach significance after FDR correction (P > 0.1) and the correlation after removing the three outliers is only r = 0.31 (P > 0.1). Most of this positive correlation is driven by correspondence of ERPs and α topographies, as indicated by partial correlation analysis. The partial correlation between γhigh AA and ERP rms levels, controlling for α AA, is essentially zero (r = −0.04, P > 0.1). The overall pattern suggests that ERPs and γhigh AA are not correlated except at the sites showing the strongest γhigh AA responses, which also typically show strong ERPs. These focally activated sites are suggested to participate in the specific (information processing) response, whereas the widely dispersed ERPs that show no correspondence to γhigh AA are part of the nonspecific response. The topography results also support the claim that ERPs, at least in nonprimary cortices, are composed of only lower frequencies (≤30 Hz). This is formally assessed with intertrial coherence (ITC) analyses.

The ITC is a measure of phase consistency across trials (see METHODS), where high versus low ITC corresponds to “evoked” versus “induced” activity. As suggested by the label “evoked,” regions of the TF plane with high ITC are expected to contribute to the ERPs. We observe that the lower frequencies (4–15 Hz) often exhibit significant (P < 0.05) ITC increases during the interval 50–200 ms and this low-frequency ITC is greatest in the electrodes exhibiting the largest ERPs. In Fig. 2, both the ITC and ERP amplitude (Fig. 2, C and D) increase going from left to right (from Patient 1 to 6 to 8). The ITC increase is largest in the θ band (4–7 Hz) and often the α band (7–15 Hz). However, increased ITC is only occasionally (i.e., Patient 6 in Fig. 2) found in the β band (15–30 Hz) and in no electrode is a significant increase found in the γ band (≥30 Hz). To assess the correspondence of low-frequency ITC to ERP rms across all electrodes, we take the frequency range 4–15 Hz (θ and α bands) for the ITC. In Fig. 3D, the topographical results across all electrodes for the ITC in the range 4–15 Hz are seen to exhibit a good correspondence to the ERPs (Fig. 3C). This visual impression is confirmed with the correlation analysis (Fig. 3G), where 4– to 15-Hz ITC is found to be significantly correlated to ERP rms (r = 0.69, P < 0.01). Note that the correlation is not expected to be perfect given the use of phase (vs. linear) coherence, the use of ERP rms (rather than AA in each frequency band), and the fact that lower frequencies (≠4–15 Hz) also contribute in part to ERPs but are not analyzed here. We noted earlier that 4– to 7-Hz AA in the latency range 50–200 ms is strongly correlated with ERP rms (r = 0.69, P < 0.01). Together the AA and ITC results confirm that the ERP is composed of lower frequencies (≤30 Hz), both in the single trials and in the average. Most of the observed ERPs are long-latency auditory evoked potentials (LAEPs) and are fo-

---

**TABLE 1. Number of electrodes exhibiting γ\text{high} activations in each macroanatomical region**

<table>
<thead>
<tr>
<th>Region</th>
<th>Significant</th>
<th>Nonsignificant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFG</td>
<td>0</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>iPv</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Par. Opp.</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TPI</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Par.-Occ.</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>MTG</td>
<td>0</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Mid-STG</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ant-STG</td>
<td>0</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>

Values are the number of electrodes for each region that exhibit each level of γ\text{high} activation. These are determined from the mean of the AA from 50 to 200 ms poststimulus (i.e., the same numbers used in Fig. 3A). The average γ\text{high} level over this interval was strong if >10% and weak if 1.5–5%. It was significant if >50% of the P values within this interval were significant (P < 0.05) and if the average P value over this interval was <0.1. The anatomical labels are as follows: IFG, inferior frontal gyrus; iPv, inferior portion of ventral premotor cortex (the “i” distinguishes it from iPm that is thought to give auditory responses); Par. Opp., parietal operculum; TPI, temporoparietal junction; MTG, middle temporal gyrus; mid-STG, mid superior temporal gyrus (defined as below the pre- or postcentral gyrus); ant-STG, anterior STG.
cused particularly on the θ band (4–7 Hz), where the strongest AA and ITC results are found in the ERP latency range.

In contrast, the $γ_{\text{high}}$ response is not phase consistent across trials (for all latencies and electrodes, ITC ≤0.2, $P > 0.05$) and is therefore purely “induced.” There is no significant correlation ($r = 0.13$, $P > 0.1$) across electrodes between $γ_{\text{high}}$ ITC (70–160 Hz, 50–200 ms) and ERP rms (50–200 ms) (see methods on the need to carefully control for the number of trials in this analysis). Along with the lack of robust correspondence for ERPs and $γ_{\text{high}}$ AA given earlier, this confirms that $γ_{\text{high}}$ does not significantly contribute to the ERPs.

Phoneme category differences

Although one original motivation of this study was to examine phoneme mismatch responses, conclusive results were not obtained for any of the measures ($γ_{\text{high}}$, α, or ERPs). For the $γ_{\text{high}}$ band, only 4 of 45 electrodes, all on the mid to anterior STG, exhibited appreciable differences between the /ua/ deviant and the /aa/ standard in the MMN block. Of these, 2 electrodes were from Patient 8, who did not complete the control block (for comparing the /ua/ deviants with the /ua/ controls), and one electrode did not reach significance ($P > 0.05$). Thus only one electrode exhibited a strong response to the /ua/ deviant that was significantly greater than both the /aa/ standard and the /ua/ control. This electrode was on the anterior STG, but ≈9 other electrodes of similar location showed no STA versus DEV differences. Therefore further detailed analyses are not shown; possible reasons for the negative results are discussed in the following text.

Reanalysis of tone MMN data

Our main novel finding is the correlation between an initial (50–200 ms) $γ_{\text{high}}$ increase and a subsequent (200–450 ms) α suppression. Here we confirm this finding by reanalyzing the data from our prior ECoG study in tumor patients (Edwards et al. 2005), where 500- and 550-Hz tones were presented for passive listening in an MMN study. Specific details of these methods can be found in that study, except that here we use the current TF analysis methods and settings. The SOA was considerably shorter in the tone study (424 ± 29 ms) than that in the present study (775 ± 50 ms), which is less ideal for studying α suppression (which habituates fairly rapidly). Nonetheless, the correlation between $γ_{\text{high}}$ AA (50–200 ms) and α AA (200–400 ms) was $-0.53$ ($P < 0.01$). By removing the “outlier” electrodes (those exhibiting >10% $γ_{\text{high}}$ increases), the correlation between $γ_{\text{high}}$ AA and α AA remained $-0.40$ ($P < 0.02$), confirming the finding of the present study.

Discussion

Overall TF and ERP responses

We studied the ECoG responses of exposed human cortex to simple speech stimuli during passive listening. Responses were characterized with TF analyses to obtain analytic amplitude (AA) and intertrial coherence (ITC) for a set of frequency bands centered between 4 and 250 Hz. Robust TF responses to syllable stimuli were obtained from the mid-STG. These responses exhibited a characteristic form (Ectors 1936), also observed in our prior study (Edwards et al. 2005), and are suggested to represent the “canonical” form for activated cortex (Edwards 2007). Higher frequencies (>30 Hz), particularly $γ_{\text{high}}$ (studied here from 70 to 160 Hz), show poststimulus AA increases. Lower frequencies (<30 Hz) show AA decreases, often called “event-related desynchronization” (ERD) (Pfurtscheller and Aranibar 1977). In addition to the ERD response, there is often an initial (~20–200 ms) increase in low-frequency AA associated with the ERP. Confirming that ERPs are composed of only lower frequencies, the ITC shows significant increases only for frequencies ≤30 Hz and these correlate across electrodes with ERP rms levels. In contrast, no significant ITC was observed for the $γ_{\text{high}}$ band, i.e., the $γ_{\text{high}}$ increases are purely “induced.”

The strongest responses for all measures were obtained from the mid-STG. For $γ_{\text{high}}$, strong increases (>5% relative to baseline) in AA were found only in the mid-STG with a small number of electrodes in other perisylvian sites showing weak (<5%) but significant increases (Table 1). It must be emphasized, however, that we lack coverage of primary auditory cortex and surrounding areas (superior temporal plane and posterior STG) that are also expected to respond strongly. α suppression was strongest over the STG, but was also observed with a more widespread topography. ERPs also exhibited a more widespread topography, one that was correlated with but not identical to the α topography. As in our prior study that also used passive listening, little auditory responsiveness was observed from the left IFG other than moderate α suppressions. Only two electrodes on the left IFG exhibited very weak $γ_{\text{high}}$ and ERP responses were similarly absent.

The $γ_{\text{high}}$ responses from the mid-STG are consistent with the known cortical organization for speech processing. Other intracranial studies in humans consistently implicate the mid to posterior STG in acoustic-phonetic processing of speech stimuli, including evidence from single-unit recording (Creutzfeldt et al. 1989), electrical stimulation mapping (Boatman 2004), and ECoG recordings in epilepsy patients (Canolty et al. 2007; Crone et al. 2001; Soltani 2007). Mid to posterior STG sites are also consistently implicated in positron emission tomography (PET) and fMRI studies (Démonet et al. 1992; Indefrey and Cutler 2004; Scott et al. 2000; Zatorre et al. 1996). Given that we examined only passive listening, the absence of IFC responses is also consistent with PET and fMRI studies that compared passive versus active tasks (Binder et al. 1996; Démonet et al. 1992; Fiez et al. 1995; Zatorre et al. 1996). Thus the present study adds to the growing evidence for a strong correspondence between ECoG $γ_{\text{high}}$ and local blood flow (Brovelli et al. 2005; Lachaux et al. 2007; Mukamel et al. 2005). The important implication is that the same topographical results that are obtained with PET and fMRI studies can be obtained by $γ_{\text{high}}$ mapping, but with superior temporal resolution. Scalp EEG also yields excellent temporal resolution, but its “field-of-view” is on the order of about 6 cm² of cortex (Cooper et al. 1965) compared with <1 cm² for ECoG $γ_{\text{high}}$ (this follows from observations of completely distinct $γ_{\text{high}}$ responses at 1-cm spatial separation; see for example the cited work by Crone et al.).

In contrast to the focal $γ_{\text{high}}$ topography, event-related decreases in α AA were observed at widespread locations, including STG, MTG, and IFG. This contrast confirms the same general finding in several prior ECoG studies (Canolty et al. 2007; Crone et al. 2001; Edwards et al. 2005; Miller et al. 2007).
2007) and indicates the more nonspecific character of \( \alpha \) suppression. We report here the novel finding that weak \( \gamma_{\text{high}} \) responses also occur at these widespread sites of \( \alpha \) suppression, with a significant correlation between the magnitudes of the two responses. The strongest correlation is obtained for \( \gamma_{\text{high}} \) leading \( \alpha \) suppression by about 100 ms. This indicates that a small increase in local firing rate and fast synaptic activity (\( \gamma_{\text{high}} \)) accompanies the triggering mechanism for the \( \alpha \) suppression.

Significant ERPs were obtained from multiple electrodes over STG and MTG sites and one electrode on the inferior precentral gyrus. No consistent morphology was observed, although most responses resembled at least in part the long-latency auditory evoked potentials (LAEPs, P1–N1–P2). The middle-latency auditory evoked potentials (MAEPs) were not obtained with the present coverage, except perhaps in one mid-STG electrode. As in our prior study (Edwards et al. 2005), only few or no ERP responses were obtained from left prefrontal (IFG) sites during passive listening. We discussed there the possibility that frontal auditory responses arise mainly from the right hemisphere and the present results remain consistent with this view.

**Limitations of the ECoG method**

There are two major limitations of the ECoG method, which in humans is performed almost exclusively in tumor or epilepsy patients. The first is the lack of spatial coverage, most notably of intrasulcal regions that constitute about two thirds of the cortex. The electrode placement is determined by clinical purposes, so systematic spatial coverage of regions of interest is not possible. As in animal microelectrode studies, where usually only one electrode is studied at a time, topographic results must be built up over multiple sequential recordings. In epilepsy recordings with grids of electrodes, more systematic coverage is obtained simultaneously from one individual, but still the intrasulcal regions are not sampled. The second major limitation is that ECoG is performed only in neurological patients with tumors or epilepsy. In epilepsy patients, atypical high-frequency oscillations have been observed associated with the epileptic pathology (Fisher et al. 1992; Grenier et al. 2003; Traub et al. 2001). The present study in tumor patients, where such high-frequency abnormalities are not generally observed, provides a confirmation that \( \gamma_{\text{high}} \) results obtained in epilepsy patients are a ubiquitous feature of nonepileptic human cortex. However, abnormal \( \delta \) waves have long been observed in the vicinity of tumors (Scarf and Rahm Jr 1941; Walter 1936) and it is possible that some of our electrodes are over subnormal cortex. Although we do not study the \( \delta \) frequency range (<4 Hz) here, it cannot be excluded that some cortical reorganization has occurred in the vicinity of the tumor. Both spatial undersampling and the possibility of subnormal cortical function are more likely to result in false negatives rather than false positives, so negative results should be interpreted with particular caution in human ECoG studies. Note that both of these limitations could potentially be overcome in animal ECoG and local field potential (LFP) studies, so we expect \( \gamma_{\text{high}} \) mapping to play an increasing role in future animal studies.

**Negative results for phoneme mismatch responses**

Although an original motivation of this study was to examine phoneme mismatch responses, conclusive results were not obtained for any of the measures (\( \gamma_{\text{high}} \), \( \alpha \), or ERPs). Only one electrode exhibited a \( \gamma_{\text{high}} \) response to the /uu/ deviant that was significantly greater than both the /aa/ standard and the /uu/ control. This electrode was on the anterior STG, but at least nine other electrodes of similar location showed no standard versus deviant differences. Results for ERPs and \( \alpha \) were similarly inconclusive. One possible explanation for this pattern is that the level of categorical representation in speech processing is located primarily within the superior temporal sulcus (STS) and extends onto the exposed surface of the STG in only a subset of individuals. However, without coverage of intrasulcal cortex, no definite conclusions can be drawn. Another possibility is that the stimuli are not ideal for studying this question, since the standard and deviant are both vowels that are not strongly contrasting. Perhaps stronger results would have been obtained for stronger categorical contrasts of consonant stimuli (e.g., voiced vs. voiceless, etc.).

**Contrasting origins of ERPs and \( \gamma_{\text{high}} \)**

It might be argued that \( \gamma_{\text{high}} \) represents a mathematical artifact of the sharp rising or falling phase of ERPs. A recent mathematical modeling study indicates this as a possibility in some circumstances (Kramer et al. 2008), but we conclude against this possibility for several reasons. First, \( \gamma_{\text{high}} \) activations do not necessarily peak at the rising or falling phases of ERPs and often persist beyond the latency of ERPs. Second, ERPs and \( \gamma_{\text{high}} \) occur at topographically distinct locations and ERPs are not necessarily sharper at electrodes exhibiting \( \gamma_{\text{high}} \). We found little correlation between the ERPs and \( \gamma_{\text{high}} \) responses across electrodes (Fig. 3E). Third, ERPs are not visibly changed in their rising, peak, or falling phases by filtering <50 Hz (except small contributions from 60-Hz line noise). Finally, \( \gamma_{\text{high}} \) increases do not show significant ITC, as would be expected if they contributed to any part of the evoked response (see also Steinschneider et al. 2008).

ERPs and \( \gamma_{\text{high}} \) must be considered to be distinct phenomena, both mathematically and in terms their physiological/biophysical origins. Studies of EEG and LFP correlations with cerebral blood flow consistently show a robust positive correlation between blood flow and \( \gamma_{\text{high}} \) activity (Brovelli et al. 2005; Darrow and Graf 1945; Lachaux et al. 2007; Logothetis et al. 2001; Mukamel et al. 2005; Niessing et al. 2005; Nir et al. 2007), whereas correspondence between ERPs and blood flow is not always found (Brovelli et al. 2005; Foucher et al. 2003). We conclude by outlining a model to account for the observed differences between ERPs and \( \gamma_{\text{high}} \). This discussion mostly concerns longer-latency secondary ERPs as observed here, as opposed to the early-latency primary ERPs (Marshall et al. 1941) that can reflect specific processing and correlate strongly with blood flow (Brinker et al. 1999).

The EEG in the frequency range contributing to ERPs (~1–30 Hz) has long been considered to originate from summed postsynaptic potentials (PSPs) on cortical pyramidal cells (Eccles 1951; Elul 1971; Purpura 1959). Excitatory postsynaptic potentials (EPSPs) on the soma or basal dendrites yield surface-positive ERPs, whereas EPSPs in apical dendrites...
yield surface-negative ERPs (Creutzfeldt and Houchin 1974; Holmes and Houchin 1967; Mitzdorf 1985). Inhibitory postsynaptic potentials (IPSPs) yield the opposite pattern, so there is no unambiguous interpretation of surface-positive versus -negative potentials. Overall, ERPs can be considered to represent changes in the balance of excitatory and inhibitory synaptic input to the basal versus apical compartments of pyramidal cells. These inputs can arrive via specific or nonspecific thalamic inputs, feedback or feedback cortical inputs, or from local intracranial input, so definite conclusions can rarely be drawn.

The EEG in the $\gamma_{\text{high}}$ frequency range (≈60–300 Hz) was not traditionally studied (but see Craggs 1974; Ectors 1936) and was only recently studied in humans by Crone and colleagues (1998, 2001). In our earlier study (Edwards et al. 2005), we discussed the diverse set of high-frequency phenomena observed in different EEG and LFP studies. It was concluded that all phenomena above the $\gamma_{\text{high}}$ range (>300 Hz) are due primarily to summed activation potential (AP) currents, which are the fastest ionic currents in cortex. $\gamma_{\text{high}}$ was argued to represent the summation of fast EPSP and IPSP currents in addition to AP currents (the same conclusion was reached on the basis of multilaminar recordings by Steinwender et al. 2008). Primarily $\alpha$-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)–mediated EPSPs and $\gamma$-aminobutyric acid type A (GABA$_A$)–mediated IPSPs are predicted to contribute to $\gamma_{\text{high}}$ frequencies, whereas slower receptor types (i.e., $N$-methyl-D-aspartate [NMDA] and GABA$_B$) are not expected to contribute significantly. The kinetics of individual EPSPs and IPSPs can be modeled for different receptor types (Destexhe et al. 1994, 1995) and the summation of many thousands of such events in the extracellular field can then be examined. By a simple statistical–mechanical argument (Edwards 2007; Motozawa 1943), it can be shown that an increase in the number of fast PSP events will lead, with or without synchrony, to an increase in the population-summed level of $\gamma_{\text{high}}$. Based on estimates of the energy budget for gray matter signaling, an increase in the number of these ionic current events (APs, GABA$_A$–IPSPs, and AMPA–EPSPs) will also lead to a strong increase in local metabolic usage to restore ionic balances (Ames 3rd 2000; Attwell and Laughlin 2001; Lennie 2003; Roland 1993). Noting the nearly linear relation between blood flow and metabolism (Gerard 1938; Hoge et al. 1999; Roland 1993), this gives a straightforward explanation for the consistent observations of positive correlations between $\gamma_{\text{high}}$ levels and blood flow. In sum, the $\gamma_{\text{high}}$ envelope represents the envelope of the fastest and most metabolically expensive ionic current activities.

Based on these considerations, the observed differences between ERPs and $\gamma_{\text{high}}$ may be better understood. Changes in the balance of synaptic inputs across cortical layers may or may not result in local activation, in the sense of increased firing rates, metabolic usage, and blood flow. Thus the presence of an ERP does not guarantee an increase in $\gamma_{\text{high}}$. However, increased local processing generally implies a change of inputs involving a transient perturbation of synaptic balances across layers. Thus sites exhibiting $\gamma_{\text{high}}$ generally also show a significant ERP. Whereas $\gamma_{\text{high}}$ can be explained as a nonsynchronous increase in the number of fast ionic currents, ERPs are best explained as the coincidence (synchrony) in time and space (apical vs. basal) of faster and slower ionic currents.

Whereas ERPs are just as likely to result from diffuse, nonspecific inputs to cortex, strong $\gamma_{\text{high}}$ increases result primarily from specific, focal activation of cortex.

ACKNOWLEDGMENTS

We thank R. T. Canolty, C. Clayworth, and R. Fechter for assistance.

GRANTS

This work was supported by National Institute of Neurological Disorders and Stroke (NINDS) Grant NS-21135, NINDS Fellowship JF32-NS-06166-01, and National Institute on Deafness and Other Communication Disorders Grants DC-006435 and DC-4855.

REFERENCES


J Neurophysiol · VOL 102 · JULY 2009 · www.jn.org