Sound identification in human auditory cortex: Differential contribution of local field potentials and high gamma power as revealed by direct intracranial recordings

Kirill V. Nourski[1,*], **Mitchell Steinschneider**[2,*], **Ariane E. Rhone**[1], **Hiroyuki Oya**[1], **Hiroto Kawasaki**[1], **Matthew A. Howard III**[1], and **Bob McMurray**[3]

[1]Department of Neurosurgery, The University of Iowa, Iowa City, IA, 52242 USA

[2]Department of Neurology, Department of Neuroscience, Albert Einstein College of Medicine, New York, NY 10461, USA

[3]Department of Psychology, Department of Communication Sciences and Disorders, Department of Linguistics, The University of Iowa, Iowa City, IA, 52242 USA

Abstract

High gamma power has become the principal means of assessing auditory cortical activation in human intracranial studies, albeit at the expense of low frequency local field potentials (LFPs). It is unclear whether limiting analyses to high gamma impedes ability of clarifying auditory cortical organization. We compared the two measures obtained from posterolateral superior temporal gyrus (PLST) and evaluated their relative utility in sound categorization. Subjects were neurosurgical patients undergoing invasive monitoring for medically refractory epilepsy. Stimuli (consonant-vowel syllables varying in voicing and place of articulation and control tones) elicited robust evoked potentials and high gamma activity on PLST. LFPs had greater across-subject variability, yet yielded higher classification accuracy, relative to high gamma power. Classification was enhanced by including temporal detail of LFPs and combining LFP and high gamma. We conclude that future studies should consider utilizing both LFP and high gamma when investigating the functional organization of human auditory cortex.

Keywords

Averaged evoked potential; Classification analysis; Electrocorticography; Speech; Superior temporal gyrus

Corresponding author: Kirill V. Nourski, MD, PhD, Department of Neurosurgery, The University of Iowa, 200 Hawkins Dr 1815 JCP, Iowa City, IA 52242 USA, Phone: +1 (319) 335-7049, kirill-nourski@uiowa.edu.
*These authors contributed equally to this study

Introduction

Intracranial recordings (electrocorticography or ECoG) have become crucial for identifying the functional organization of human auditory cortex due to their high spatial and temporal resolution (e.g. Mukamel and Fried, 2012; Nourski and Howard, 2015). ECoG is a rich time-varying measure that simultaneously reflects synaptic activity and action potentials from populations of neurons. Consequently, there is considerable interest in exactly which aspects of the ECoG signal recorded from the auditory cortex carry information relevant to sound processing. Addressing this is not only important methodologically but may also hint at fundamental ways in which populations of neurons code information.

Earlier studies using the intracranial methodology relied on analysis of time domain-averaged local field potential (LFP) signals to examine response properties of auditory cortex (e.g. Celesia and Puletti, 1969; Lee et al., 1984; Liégeois-Chauvel et al., 1991, 1994; Halgren et al., 1995; Steinschneider et al., 1999; Howard et al., 2000). The averaged LFP (i.e., averaged evoked potential, AEP) emphasizes relatively low-frequency components of the ECoG signal that are both time- and phase-locked to the stimulus. This approach was in part aimed at identifying cortical generators of specific components of AEPs recorded using electroencephalographic methods and neuromagnetic fields recorded using magnetoencephalography in response to sound stimuli (e.g. Liégeois-Chauvel et al., 1994; Halgren et al., 1995; Howard et al., 2000).

With few exceptions (e.g., Sahin et al., 2009; Sinai et al., 2009; Chang et al., 2010; Steinschneider et al., 2011), more recent studies using intracranial methodology have focused on event-related band power (ERBP) of the high gamma frequency component (70–150 Hz) of the ECoG (e.g. Crone et al., 2001). This approach often occurred at the expense of analysis of the time- and phase-locked activity as captured by the LFP. This paradigm shift has been driven by the findings that enhanced high gamma activity in the ECoG is closely related to increases in the blood oxygenation level-dependent signal as measured by functional magnetic resonance imaging methodology and to spiking activity in cortical neurons (Mukamel et al., 2005; Nir et al., 2007; Steinschneider et al., 2008). The focus on high gamma activity has yielded new understandings of the functional organization of human auditory and auditory-related cortex over the last decade. For instance, this analysis has helped characterize the functional representation of phonetic categories used in speech (Pasley et al., 2011; Mesgarani et al, 2014) and the powerful effects of selective attention in modulating auditory cortical activity when listening to competing speech streams (e.g., Mesgarani and Chang, 2012). Analysis of high gamma activity has also demonstrated tiered effects of attention across auditory and auditory-related cortical areas (Nourski et al., 2014b; Steinschneider et al., 2014). All these findings parallel those obtained from patterns of spiking activity in experimental animals (e.g., Fritz et al., 2003; Mesgarani et al., 2008; Tsunada et al., 2011; Steinschneider et al., 2013). However, at the same time, the aforementioned advances have not compared the utility of LFPs and high gamma activity in understanding auditory processing.

This focus may be limiting because intracranially recorded LFPs have also helped characterize underlying features of functional organization of auditory and auditory-related

cortex (e.g. Brugge et al., 2008; Chang et al., 2010). Furthermore, those studies that examined both high gamma and LFP revealed differences in the ways the two metrics are related to stimulus acoustics and perception (e.g., Nourski et al., 2009; Nourski and Brugge, 2011). These studies raise the possibility that both the LFP and high gamma ERBP in the ECoG reflect relevant (and non-redundant) information about sounds. If this is the case, it raises key questions about what information may be carried by the LFP that is not seen in the high gamma activity. What is needed is a direct comparison of the two measures of cortical activity to determine their relative contributions for carrying meaningful information about complex auditory stimuli as a whole.

In this study, we objectively examined this issue by using a classification approach. Subjects passively listened to consonant-vowel (CV) syllables and pure tone stimuli. The contribution of different measures of cortical activity for carrying information about these stimuli was assessed by training a classifier (support vector machine, SVM) to identify properties of the stimulus [voicing, place of articulation (POA), and tone frequency] on the basis of various permutations of the LFP and high gamma ERBP signal.

Under typical preparations, the LFP is usually a linear scaled voltage signal. It can be both positive and negative. In contrast, high gamma activity is usually a rectified power signal that is logarithmically scaled and baseline-normalized. Classification analysis can abstract across the differences between LFP and high gamma signal representation, as its dependent variable is classification accuracy (in percent correct) rather than a difference in the signal *per se*. Moreover, a non-parametric approach like an SVM may be better equipped for factoring out these differences than parametric approaches. It is important to note that the LFP signal is not orthogonal or independent of the high gamma signal, as they both are derived from the same underlying ECoG waveform. Our goal was not to try to parse out the unique information contained in each signal, but rather to use these coarse measures to ask if there is anything that cognitive neuroscience may be missing by relying on one over the other. To achieve this goal, we focused on typical preparations of these signals, characterizing the LFP as voltage time series, and high gamma activity as rectified, log-transformed and baseline-normalized ERBP.

Contrary to expectations, we found that classification accuracy based on LFPs was superior to that provided by high gamma activity. Best accuracy was often obtained when both measures were included in the analysis. Methodologically, this suggests that future studies should utilize neural activity captured both by the LFP and high frequency ERBP when investigating the functional organization of human auditory and auditory-related cortex; more broadly, it raises the possibility that by studying only local high gamma power we may be missing relevant aspects of sound encoding within the auditory cortex.

Methods

Subjects

Subjects consisted of 21 neurosurgical patient volunteers (16 male, 5 female, age 20–56 years old, median age 33 years old). The subjects had medically refractory epilepsy and were undergoing chronic invasive ECoG monitoring to identify potentially resectable

seizure foci. Research protocols were approved by the University of Iowa Institutional Review Board and by the National Institutes of Health. Written informed consent was obtained from each subject. Participation in the research protocol did not interfere with acquisition of clinically required data. Subjects could rescind consent at any time without interrupting their clinical evaluation.

The patients were typically weaned from their antiepileptic medications during the monitoring period at the discretion of their treating neurologist. Experimental sessions were suspended for at least three hours if a seizure occurred, and the patient had to be alert and willing to participate for the research activities to resume.

In all participants, ECoG recordings were made from only a single hemisphere. All subjects but two had left-hemisphere language dominance, as determined by intracarotid amytal (Wada) test results; subject R149 had bilateral language dominance, and R139 had right language dominance. In ten subjects, the electrodes were implanted on the left side, while in eleven others recordings were from the right hemisphere (the side of implantation is indicated by the prefix of the subject code: L for left, R for right). The hemisphere of recording was language-dominant in twelve subjects (all subjects with left hemisphere implanted, R139, and R149) and non-dominant in nine other subjects (R127, R129, R136, R142, R175, R180, R186, R210 and R212). All subjects were native English speakers. Intracranial recordings revealed that the auditory cortical areas on the superior temporal gyrus were not epileptic foci in any of the subjects.

Subjects underwent audiometric and neuropsychological evaluation before the study, and none were found to have hearing or cognitive deficits that could impact the findings presented in this study. Sixteen out of twenty-one subjects had pure-tone thresholds within 25 dB HL between 250 and 4 kHz. Subjects R127, L130 and R212 had a mild (30–35 dB HL) notch at 4 kHz, subject L258 had bilateral mild (25–30 dB HL) low frequency hearing loss at 125–250 Hz, and subject L145 had bilateral moderate (50–60 dB HL) high frequency hearing loss at 4–8 kHz. Word recognition scores, as evaluated by spondees presented via monitored live voice, were 100%, 98% and >91% in 13, 4, and 2 subjects, respectively. Speech reception thresholds were better than 20 dB in all tested subjects, including those with tone audiometry thresholds outside the 25 dB HL criterion. Importantly, frication and formant transition information relevant for consonant identification in the experimental stimuli (see below) was at or below 3 kHz, thus mitigating the detrimental effects on speech perception in subjects with hearing loss at 4 kHz and above.

Stimuli

Experimental stimuli were synthesized stop consonant-vowel (CV) syllables, used previously in Steinschneider et al. (1999, 2005, 2011), and pure tones. The speech syllables /ba/, /da/, /ga/, /pa/, /ta/ and /ka/ were constructed on the cascade branch of a KLSYN88a speech synthesizer (Klatt and Klatt, 1990). They contained 4 formants (F1 through F4), and were 175 ms in duration.

Synthesis was based on the voiced (/ba/, /da/, /ga/) tokens. For these, fundamental frequency began at 120 Hz and fell linearly to 80 Hz. All syllables had the same F1, which had an

onset frequency of 200 Hz and rose over a 30 ms transition to a 700 Hz steady state. F2 started at 800 Hz for /b/, and 1600 Hz for /d/ and /g/; it reached its steady state of 1200 Hz after a 40 ms transition. F3 started at 2000 Hz for /b/, 3000 Hz for /d/ or 2000 Hz for /g/ and reached a steady state of 2500 Hz after a 40 ms transition. F4 did not contain a transition and was flat at 3600 Hz. Thus, /ga/ had the same F2 as /da/ and the same F3 as /ba/. A 5 ms period of frication exciting F2–F4 preceded the onset of voicing. Amplitude of frication was increased by 18 dB at the start of F2 for /ba/ and /ga/, and F3 for /da/. The result of this synthesis was that /ba/ and /da/ had diffuse onset spectra maximal at low or high frequencies, respectively, whereas /ga/ had a compact onset spectrum maximal at intermediate values, consistent with Stevens & Blumstein (1978). The voiceless syllables (/ pa/, /ta/, and /ka/) were identical to their voiced counterparts (/ba/, /da/, and /ga/) except for an increase in the voice onset time (VOT) from 5 to 40 ms. This was done by eliminating the amplitude of voicing for the 35 msec after the frication and replacing this with aspiration. In 10 subjects out of 21, a control set of pure tone stimuli was also used. The tones had frequencies matching the 18-dB boosted formants of the CV syllables (800, 1600 and 3000 Hz) and were 175 ms in duration.

Stimuli were delivered to both ears via insert earphones (ER4B, Etymotic Research, Elk Grove Village, IL) that were integrated into custom-fit earmolds. Stimuli were presented at a comfortable level (mean = 65 dB SPL, SD = 5.7 dB SPL). Inter-stimulus interval was chosen randomly within a Gaussian distribution (mean interval 2 s; SD = 10 ms) to reduce heterodyning in the recordings secondary to power line noise. Stimulus delivery and data acquisition were controlled by a TDT RP2.1 and RX5 or RZ2 real-time processor (Tucker-Davis Technologies, Alachua, FL).

Design and Procedure

Experiments were carried out in a dedicated electrically-shielded suite in The University of Iowa General Clinical Research Center. The room was quiet, with lights dimmed. Subjects were awake and reclining in a hospital bed or an armchair. Stimuli were randomized and presented in a passive listening paradigm, without any task direction. There were 50 repetitions of each stimulus.

Recording

Recordings were made from perisylvian cortex including the superior temporal gyrus (STG) (Howard et al., 2000, Nourski and Howard, 2015) using high density subdural grid electrodes (AdTech, Racine, WI). In all subjects but one (R149), the recording arrays consisted of 96 platinum-iridium disc electrodes (2.3 mm exposed diameter, 5 mm center-to-center spacing) arranged in an 8×12 grid and embedded in a silicon membrane. In subject R149, an 8×8 grid (64 recording electrodes) was used. A subgaleal contact was used as a reference. Recording electrodes remained in place for approximately 2 weeks under the direction of the patients' neurologists. Anatomical locations of recording sites were determined using structural magnetic resonance imaging, high-resolution computed tomography, and intraoperative photography. Data from recording sites overlying posterolateral superior temporal gyrus (PLST) were included in the analysis.

Data processing

Recorded ECoG data were filtered (1.6–1000 Hz bandpass, 12 dB/octave rolloff), amplified (20×), and digitized at a sampling rate of 2034.5 Hz. Data analysis was performed using custom software written in MATLAB Version 7.14.0 (MathWorks, Natick, MA, USA). Pre-processing of ECoG data included downsampling to 1 kHz for computational efficiency, followed by removal of power line noise by an adaptive notch filtering procedure (Nourski et al., 2013). Evoked cortical activity was visualized as AEPs, calculated by across-trial averaging of ECoG waveforms. Analysis of ERBP focused on the high gamma ECoG frequency band. High gamma ERBP was calculated for each recording site. Single-trial ECoG waveforms were bandpass filtered between 70 and 150 Hz (300th order finite impulse response filter), followed by Hilbert envelope extraction. The resultant high gamma power envelope waveforms were log-transformed, normalized to power in a prestimulus reference window (250-50 ms prior to stimulus onset), and averaged across trials for visualization. Peak-to-peak AEP amplitudes and peak high gamma ERBP were measured for each recording site overlying PLST within a time window of 0–200 ms after stimulus onset. Comparisons of AEP and high gamma response magnitudes between left and right hemisphere recording sites were made by averaging across sites and using independent-samples t-tests.

Classification analysis

Classification analysis examined the relative amount of information contained in distributed patterns of activity across STG using either the local field potential (LFP) waveform or high gamma ERBP captured from individual trials. To accomplish this, a combination of LFP and/or high gamma ERBP at each STG site at each trial was used as a set of features to train a SVM classifier. These signals were extracted from 50 ms-wide time windows centered every 25 ms from 200 ms before stimulus onset to 1200 ms after stimulus onset. SVMs were trained to predict from the pattern of activity distributed across recording sites one of three properties of the stimulus: 1) the voicing of the speech sound (discriminating /b, d, g/ from /p, t, k/); 2) POA of that speech sound (a three-way contrast between /b/, /d/ and /g/); or 3) the frequency of a pure tone (800, 1600 or 3000 Hz). For the POA task, the analysis was also replicated with the three voiceless syllables /p, t, k/ and with all six syllables (i.e., a three-way contrast between /b, p/, /d, t/ and /g, k/), and highly similar results were obtained.

SVMs were trained separately for each subject for each task (voicing, POA, tone frequency) with a 15-fold procedure. Training data was carved into 15 approximately equal samples, each containing approximately the same number of trials from each of the stimuli to be discriminated. The SVM was then trained to identify boundaries on the basis of 14/15ths of the data and tested on the remaining 15th. Analysis then cycled through each 15th of the data until testing data was complete for the entire stimulus set. This was then repeated 10 times (to account for variations in the random partitioning of the set) to obtain a final mean accuracy of a given classification task.

For the voicing classification there were 150 voiced sounds (/b, d, g/) and 150 voiceless sounds (/p, t, k/); when submitted to the 15-fold classification scheme this led to 280 training tokens and 20 testing tokens (per fold). For the place of articulation classification, there

were 50 tokens of /b/, 50 of /d/ and 50 of /g/; this led to 140 tokens used for training and 10 for test (on each fold). Tone classification had the same data structure as POA.

SVMs were implemented in SVMlib (Chang & Lin, 2011). SVMs used the radial basis function as a kernel. Potential features (independent variables) for the SVMs included the LFP and high gamma ERBP (distributed across recording sites), averaged within 50 ms time windows. To account for information present in the temporal dynamics of each signal within that time window, the linear slope of the LFP or high gamma over time as well as their quadratic effects were estimated for use as potential features. This led to a 4×4 design matrix, wherein each measure (LFP or ERBP) could be (1) not present, (2) estimated as a mean; (3) mean +slope or (4) mean +slope + quadratic, thus yielding a total of 15 input combinations for each classification task (if neither measure was present, no analysis was possible).

SVMs have two free parameters, the penalty or cost of incorrect categorization, C, and the width of the kernel, Γ. For each subject, for each classification task (place, voicing, tone frequency), and for each particular combination of LFP and high gamma ERBP, the optimal values for C and Γ were estimated once using brute force search using all combinations of 8 possible Cs (2^5 to 2^{19}) and 7 possible Γs (2^{-5} to 2^{-17}), logarithmically spaced using the data from a 50 ms window centered at 100 ms after stimulus onset. Consequently, the analyses for the high gamma used Cs and Γs that were optimal for high gamma; analyses for LFPs used independently estimated Cs and Γs that were optimal for the LFP (and so forth for each of the 15 combinations). We were concerned that with so many classification jobs for different values of C/Γ, this could inflate the accuracy scores, since randomness across the K-folds would naturally lead to variation from run to run. To guard against this, after the optimal C and Γ were estimated, all classification jobs were redone with new randomizations of the data. From this final set of classifications, accuracy from the SVM was computed as the proportion of correctly identified tokens. These were analyzed using ordinary ANOVA and scaled with the empirical logit function (so that they distribute more linearly). Prior to conducting the ANOVAs, the data were checked for sphericity with Mauchly's test of sphericity. When violations were found, degrees of freedom were adjusted with the Greenhouse-Geisser adjustment. These are indicated as F_{GG}.

Results

Spatiotemporal properties of AEP and high gamma activity

CV syllables elicited robust evoked potentials and high gamma activity on PLST in both dominant and non-dominant hemispheres (Figs. 2, 3). Responses recorded from individual cortical sites appeared to be more distinct when comparing voiced and voiceless syllables (e.g. /ba, pa/) relative to the POA contrast (e.g. /ba, da/) (see Figs. 2B, 3B). For most of the subjects (of which the two shown here are representative) and for many recording sites, we found that both the initial positive and later negative peaks of the AEP were shifted to longer latencies in response to /pa/ relative to /ba/ and /da/. These differences likely reflect the 5 ms VOT for /ba/ and /da/, compared to the 40 ms VOT for /pa/. Differences across POA were much more subtle, likely reflecting the highly overlapping formants and formant transitions in these synthetic syllables (see Fig. 1).

Both positive and negative voltage deflections in the AEP and high gamma power were mainly confined to PLST. This can be seen in Figures 2c and 3c, which depict the evolution of both the AEP and high gamma ERBP over time from 50 ms to 250 ms in 50 ms steps. Clear differences between responses to voiced and voiceless stimuli can be observed as early as 100–150 ms in the neural patterns across PLST for these two representative subjects. For instance, in subject L178 the development of the initial negative wave of the AEP over PLST was present in the 100–150 ms interval for /ba/, but not for /pa/ (Fig. 2C). Similarly, the magnitude of the high gamma response in the same interval was greater for the voiced consonant relative to the voiceless (Fig. 2D). No obvious differences were observed to distinguish POA, as discussed above.

Variability of AEP and high gamma ERBP

There was considerable intra- and inter-subject variability in the timing of responses that was primarily observed in the AEP. This variability is exemplified by Figure 3, which illustrates response waveforms obtained from the non-dominant (right) hemisphere in subject R180. Intra-subject variability of the AEP is evident by the different morphology of waveforms recorded from sites X and Y. Inter-subject variability can be observed in the timing of the positive and negative peaks in the AEP, which was earlier than that seen in the dominant (left) hemisphere of subject L178 (cf. Fig. 2B). In contrast, the timing of high gamma activity was comparable between subjects, and the principal intra-subject variability was based on responses to the speech sounds differing in their voicing.

Differences in response morphology of the AEP observed between the two exemplary subjects were representative of the variability seen across all subjects (Fig. 4). While AEPs at individual sites may display the characteristic morphology as described by Howard et al. (2000) using brief click-train stimuli, this morphology was not always preserved when the responses were averaged across the portion of the STG covered by the recording grid. In contrast, high gamma ERBP averaged across STG sites in each subject, while varying somewhat in latency, had a much more consistent time course across subjects (Fig. 4, right column).

The inter-subject variability of AEPs and high gamma responses was further examined by computing the grand average waveforms for all left (367 sites in 10 subjects) and right (356 sites in 11 subjects) hemisphere recording locations on the STG. Results are shown in Figure 5, with averaged left and right hemisphere responses plotted in teal and purple, respectively. For the AEP, and despite the pronounced inter-subject variability (see Fig. 4A), grand average waveforms featured the characteristic P_α-N_α-P_β-N_β morphology as described previously (Howard et al., 2000) (see Fig. 5A, upper panel). Peaks P_α, N_α and P_β of these grand average AEP waveforms had latencies of 56, 127 and 216 ms for the left hemisphere sites and 59, 127 and 217 ms for the right hemisphere sites. These values are later than the mean peak latencies previously reported for AEPs evoked by 100 Hz click trains (Howard et al., 2000; 45, 90 and 162 ms, respectively). The peaks of grand average high gamma responses overlapped the peak and falling phase of the N_α AEP component, consistent with previous studies (Crone et al. 2001, Steinschneider et al., 2011). Variability across subjects and recording sites, depicted as 95% confidence intervals, followed a similar time course for

the AEP and high gamma ERBP and paralleled the magnitude of the responses themselves, peaking shortly after 100 ms.

Hemispheric differences were also observed, and were characterized by larger grand average AEP peak-to-peak amplitude ($t(19)=2.19$, $p = 0.041$) and peak high gamma ERBP [$t(19) = 2.73$; $p = 0.013$)] over the right hemisphere as compared to the left hemisphere sites (Fig. 5B). These differences were observed despite no obvious systematic bias in electrode coverage between the hemispheres.

Classification analysis

Time course of classification performance—Initial classification analysis was intended to determine the optimal post-stimulus time window for each LFP and high gamma ERBP feature in each classification task. The goal at this point was simply to determine a single time window at which to conduct more detailed analyses of the different possible feature-sets. For this analysis, the slope and quadratic effect of LFP and high gamma were ignored, and SVM classifiers were trained with mean LFP ($\overline{V}$), mean high gamma ERBP ($\overline{dB}$), or both of them. The data were extracted from sliding 50 ms windows (25 ms overlap) centered from −0.2 to +1.2 s after stimulus onset.

Results of this analysis are shown in Figure 6A for voicing, POA and tone classification. Voicing classification accuracy was excellent, peaking at 84.56%, and was superior to tone (60.10%) and POA (44.63%) classification. While this may be attributed to the fact that the latter two had a lower chance level (~33%) than the voicing classification (50%), converting accuracy to odds ratios indicates that voicing was still substantially better than POA, though similar to tones (Voicing: 1.69, Tone: 1.80, POA: 1.29). The lower classification accuracy for POA likely reflects the overlapping formant structures of these synthetic syllables, leading to speech sounds that are less distinct from one another than a set of syllables that vary in their voicing. In all three cases, LFP and LFP+high gamma classifiers outperformed high gamma alone. Finally, for all three types of classifier inputs and for all three classification tasks, performance was maximal or near-maximal at around 100 ms (see Fig. 6A). Thus, for ease of exposition, the following analyses focused on the 75–125 ms time interval. Despite the differences across features and classification goals, all of the classification tasks were significantly above chance at this 100 ms window (Table 1).

LFP vs. high gamma—The next analysis examined classifier performance at 75–125 ms after stimulus onset. Three feature sets were compared: mean LFP in that window, mean high gamma ERBP in that window, and the combination of the two. Results are shown in Figure 6B for voicing, POA and tone classification tasks. Each of the three classification jobs was analyzed with a separate one-way ANOVA comparing empirical logit-transformed accuracy as a function of measure type, and with planned comparisons comparing mean LFP to mean high gamma ERBP and their combination to mean LFP alone. For voicing, there was a significant effect of measure [$F_{GG}(1.1,22.3) = 18.48$, $p < 0.0001$]. Follow-up comparisons revealed better performance for LFP than high gamma ERBP [$F(1,20) = 12.1$, $p = 0.002$] and no additional benefit for adding high gamma to LFP [F(1,20) = 2.2, p = 0.152]. For POA, there was an overall significant effect [$F_{GG}(1.5,29.0) = 6.13$, $p = 0.005$].

Follow-up contrasts showed greater accuracy for LFP than high gamma ERBP [$F(1,20)$ =7.14, p= 0.015], and no advantage of adding both measures over LFP alone ($F(1,20)$=1.35, p=0.25). Finally, for the tone contrast, there was an overall significant effect of measure [$F_{GG}(1.1,10.0) = 6.09$, $p = 0.01$]. This time, however, the benefit for LFP over high gamma was only marginally significant [$F(1,9) = 4.42$, p= 0.065], but the addition of high gamma power to LFP was not significant [$F(1,9) = 2.12$, $p = 0.18$]. Thus, when only the mean LFP or high gamma ERBP was considered, the mean LFP was significantly (for voicing and POA) or marginally (for tone) superior to high gamma ERBP alone. There was no statistical support for added benefit of combining the two measures.

It is possible that the enhanced classification accuracy achieved by LFP was in part due to the use of both positive and negative voltage of the LFP signal, whereas high gamma activity was calculated using a rectified signal (power). The latter was a necessary processing step due to the fact that induced high gamma activity is time-locked, but not phase-locked. To address this possible confound, we also examined classification accuracy based on full-wave rectified LFP signals. Classification accuracy for the voicing task was the highest when the LFP was not rectified. A similar result was obtained for POA classification. For both tasks, high gamma power by itself yielded the poorest performance, though the rectified LFP yielded only slightly better classification accuracy (data not shown).

Another potential confound is the use of the LFP that was not low-pass filtered, and thus would include high gamma activity. This additional activity, coupled to the lower-frequency components that dominate the LFP, might have enhanced classification accuracy for the LFP, as it contained both high and low frequency components. Therefore, we performed a control analysis using LFPs that had been lowpass-filtered at 70 Hz (i.e., the lower boundary of the high gamma frequency band). For both voicing and POA, lowpass-filtered LFPs showed accuracy profiles that were almost identical to the unfiltered LFPs, and were markedly higher than those derived from high gamma power alone (data not shown).

Contribution of fine-grained temporal dynamics—While the foregoing analysis favored LFP over high gamma as a neural information carrier about speech sounds and tones, it is possible that a more precise characterization of how a given measure changes within a given time window might provide more information about the stimulus. To examine this, the slope and the quadratic effect of time (on the LFP or high gamma ERBP) were analyzed in addition to the means of those measures. This analysis addressed two issues: (1) whether these more fine-grained measures of LFP showed any improvement over the mean LFP alone, and whether adding the mean high gamma to these LFP measures would improve classification performance; and (2) whether these more fine-grained measures of high gamma ERBP provided better classification accuracy, and whether adding mean LFP offered any additional improvement.

Contribution of fine-grained temporal dynamics of the LFP—Classification accuracy on each of the three tasks is depicted in Figure 7A as a function of increasing temporal detail of the LFP measure (i.e., adding slope and quadratic effect within 50 ms windows), and of adding mean high gamma ERBP. There was a substantially greater benefit

for adding the slope or quadratic to the mean LFP. However, there was little additional benefit from adding mean high gamma ERBP. This was confirmed in a series of 3 (LFP measure: mean/slope/quadratic) × 2 (mean high gamma: yes/no) ANOVAs for each of the three classification tasks.

The first ANOVA examined voicing classification (see Fig. 7A, left panel). There was a significant effect of LFP measure [$F(2,40) = 34.38$, $p < 0.001$]. Follow-up tests suggested that this was due to the fact that adding slope to the mean LFP resulted in a significant increase in accuracy [$F(1,20) = 76.55$, $p < 0.001$], although adding the quadratic resulted in no additional benefit ($F < 1$). There was no main effect of adding mean high gamma ERBP ($F < 1$) and no interaction [$F(2,40) = 1.62$, $p = 0.21$]. Thus, while a more detailed characterization of the LFP signal within 50 ms windows (slope) could improve classification performance, adding mean high gamma ERBP did not provide further improvement.

Analysis of classifier performance on POA revealed a similar pattern of effects to voicing. There was a significant effect of LFP measure [$F_{GG}(2.0,39.0) = 24.75$, $p < 0.001$] (see Fig. 7A, middle panel). As before, planned comparisons suggested that adding the slope of LFP to the mean improved performance [$F(1,20) = 20.15$, $p < 0.001$], and that adding the quadratic led to further improvements in classification accuracy over the slope [$F(1,20) = 7.19$, $p = 0.014$]. In contrast to voicing classification, there was a significant main effect of adding mean high gamma ERBP to the LFP [$F_{GG}(1,20) = 8.29$, $p = 0.009$], but no interaction between LFP and high gamma ERBP [$F_{GG} < 1$]. Thus, adding high gamma appears to benefit performance, although as Figure 7A shows this is numerically much smaller than increasing the temporal detail of the LFP measure.

Finally, analysis of tone classification revealed a main effect of LFP measure [$F(2,18) = 13.75$, $p < 0.001$] (see Fig. 7A, right panel). This was due to a significant increase in performance when slope of LFP was added to the mean alone [$F(1,9) = 36.125$, $p < 0.0001$]. There was also a marginal additional benefit of adding the quadratic over and above the slope [$F(1,9) = 3.48$, $p = 0.09$]. In contrast to POA classification (but like voicing), the main effect of adding mean high gamma was not significant ($F < 1$) and there was no interaction between LFP measure and high gamma [$F(2,18) = 1.28$, $p = 0.30$].

Contribution of fine-grained temporal detail of high gamma power—The next set of analyses examined whether increasingly detailed measures of the time course of high gamma ERBP would improve classification accuracy. Figure 7B shows the effect of increasing specificity of the temporal dynamics of high gamma ERBP, and of adding mean LFP to voicing, POA and tone discrimination (left, middle and right panel, respectively). Adding slope and quadratic terms to mean high gamma ERBP did not significantly increase classification accuracy, and in some cases even decreased it. Further, for every measure of high gamma, adding mean LFP information provided a fairly substantial increase in performance.

This was confirmed in a series of 3 (high gamma measure: mean/slope/quadratic) × 2 (LFP mean: yes/no) ANOVAs for each of the three classification tasks. The ANOVA for voicing

found no main effect of high gamma measure [$F_{GG}(1.2, 24.7) = 4.21$, $p = 0.29$] (see Fig. 7A, left panel). Consistent with earlier findings, there was a main effect of LFP [$F_{GG}(1,20) = 34.27$, $p < 0.001$]: adding information about mean LFP significantly improved accuracy for every type of high gamma measure. There was a significant interaction [$F_{GG}(1.5,30.6) = 16.04$, $p < 0.001$], reflecting the fact that adding the LFP made a larger contribution to performance when high gamma was characterized by its mean only.

The analysis of POA showed a main effect of high gamma measure [$F_{GG}(1.6,32.6) = 4.79$, $p = 0.02$] on classification accuracy (see Fig. 7B, middle panel). However, this was due to the fact that adding increasing temporal position in the coding of the high gamma ERBP (the slope and quadratic terms) decreased performance. In fact, planned comparisons revealed no significant decrement for adding slope to the mean alone [$F(1,20) < 1$], but a significant decrement between the slope and quadratic [$F(1,20)=10.03$, $p=.005$]. As before, adding mean LFP was significant [$F_{GG}(1,20) = 11.76$, $p = 0.003$]. The interaction was marginally significant [$F_{GG}(1.4,27.4) = 3.29$, $p = 0.069$]. Thus, the only way to improve POA classification from the mean high gamma ERBP alone was to add mean LFP to it.

For tone classification, there was a main effect of high gamma measure [$F(2,18) = 12.78$, $p < 0.0001$]. As with POA, increasing temporal detail led to poorer performance. Planned comparisons revealed that adding slope to mean high gamma ERBP resulted in a significant loss of classification accuracy [$F(1,9) = 12.14$, $p = 0.007$] (Fig. 7B, right panel). Adding the quadratic resulted in a marginally significant decrement in accuracy over and above the slope [$F(1,9) = 4.39$, $p = 0.065$]. Once again, the main effect of LFP was significant [$F(1,9) = 8.07$, $p = 0.019$]: adding mean LFP to high gamma provided a gain in accuracy. High gamma and LFP did not interact [$F(2,18) = 2.08$, $p = 0.15$]. Thus, increasing temporal detail of the high gamma signal did not improve performance on tone classification, though adding mean LFP did.

The above analyses were all carried out using a 50 ms analysis window. Longer windows of integration might be more optimal for analysis of high gamma power. To address this possibility, we examined classification accuracy by analyzing high gamma responses using different time windows (50, 100, 150 and 200 ms), all centered at 100 ms after stimulus onset, and describing the temporal dynamics within each of these windows using the mean, slope and quadratic (Supplementary Fig. 1). For voicing classification analyzed using longer time windows, accuracy using mean high gamma alone decreased from 71.2% (50 ms time window) to 63.9% (200 ms time window). Adding more detailed temporal descriptors (slope and quadratic) yielded improvement in accuracy at time windows longer than 50 ms. This effect was especially prominent for the 200 ms time window, though this improvement (74%) was only somewhat better than mean ERBP in the 50 ms (71%), and still inferior to the LFP alone in the 50 ms window (77%). Results for POA analyses were more variable. In general, longer time windows yielded slightly increased accuracy when analyzing the mean alone, or when adding more precise measures of within-window temporal dynamics (slope and quadratic) (see Supplementary Fig. 1).

Interim summary—The previous analyses support the following conclusions. Increasing the temporal specificity of the LFP, but not high gamma ERBP measure, can lead to

significant gains in classification performance. This was the case for adding the slope, and in some classification tasks, the quadratic. Thus, there is information in the detailed time course of the LFP that was not available in high gamma ERBP. While adding LFP improved classification accuracy over and above high gamma alone, the converse (adding high gamma over and above LFP) was only true in isolated cases, and the effects were numerically small. This indicates that even with the relatively coarse measure of the LFP (mean+slope), there is more information carried in the LFP than in the high gamma.

The time course of information—The primary goal of the final analysis was to determine if at any time interval some version of the high gamma-based classifiers could exceed performance of the LFP-based classifiers. A secondary goal was to investigate how long the information present in either measure persisted in the cortical activity after stimulus onset. Rather than examining the accuracy over the full time course for all 15 different combinations of LFP and high gamma band power, this analysis was simplified by determining the following in each subject: (1) the single best LFP-only analysis (mean, mean+slope, or mean+slope+quadratic) across the full time course; (2) the single best high gamma-only analysis (mean, mean+slope, or mean+slope+quadratic) across the full time course; and (3) the single best analysis across all 15 combinations. The time course of classifier accuracy was then compiled for each of these three analyses. Results for each of the three classification tasks are shown in Figure 8. Three main observations can be drawn from this analysis. First, the best LFP-only classification always exceeded the best high gamma-only classification. Second, the best LFP-only classification was close (if not identical) in accuracy and time course to the best overall classification. For 29 of 52 classification analyses carried out on data from individual subjects (21 for voicing, 21 for POA, and 10 for tone), the best overall analysis did not use high gamma ERBP in any form. Third, the information present in the LFP lasted longer in time than that present in the high gamma ERBP. Thus, even when the analyses were expanded to the full time course and used features flexibly for each subject, there was substantial information in the LFP about the identity of the stimulus that was not available in the high gamma ERBP alone.

Discussion

The goal of this study was to compare the information that can be extracted by two electrophysiological measures: the LFP and the high gamma band power. These comparisons should be interpreted as a comparison of standard methods for processing the electrophysiological signal; it is important to note that the LFP and high gamma power are not orthogonal measures of neural activity. Nonetheless, a classification approach using a non-parametric classifier can overcome some of these differences to allow us to quantify the information that can be extracted from each measure.

There were two major findings from these analyses. First, there was high variability across both recording sites and subjects in the morphology of the AEP. Variability of the high gamma response was substantially smaller. Second, despite this variability, the LFP generally carried significantly more information about the identity of the syllables and tones than the high gamma activity alone. This was true even when the fine-grained temporal changes within 50 ms windows were considered for both signals over the time course of

analysis. Moreover, there was an inconsistent benefit to the use of both measures: while in some classification tasks high gamma appeared to offer additional information beyond the LFP, in others it did not. The possible causes and consequences of each finding are discussed below.

Spatiotemporal properties and variability of AEP and high gamma ERBP

As reported previously (e.g. Crone et al., 2001; Chang et al., 2010; Steinschneider et al., 2011; Nourski et al., 2014c), speech sounds as well as pure tones elicit large-amplitude AEPs and high gamma ERBP overlying PLST in both hemispheres. Ambiguity exists regarding the generators of the AEP recorded directly over the lateral surface of the STG. Several studies have suggested that Heschl's gyrus and surrounding regions (e.g., planum temporale) of the superior temporal plane provide major contributions to the AEP recorded from lateral STG (Crone et al., 2001; Edwards et al., 2005). This issue is non-trivial, as one of the major results of this study was the superiority of the low-frequency LFP in identifying test stimuli over that provided by high gamma activity. LFPs have been shown to represent both local activity within the immediate vicinity of the recording contact, as well as volume-conducted activity emanating from distant sites (Kajikawa and Schroeder, 2011). If the AEP is generated by activity within the superior temporal plane, then we would be comparing informational content from vastly different core and non-core regions of auditory cortex.

The AEPs recorded from the majority of subjects featured major peaks that showed positive-negative-positive deflections below the Sylvian fissure; a pattern consistent with radial sources generated within PLST and not tangential sources located within the superior temporal plane. This conclusion is supported by the observation that the N_1 component of the scalp-recorded AEP is maximal near the vertex, consistent with major generators within the superior temporal plane (Scherg et al., 1989; Liégeois-Chauvel et al., 1994). The timing of the N_1 component overlaps with the initial prominent negativity (N_α) observed on PLST. If the AEP recorded from PLST was generated within the superior temporal plane, one would expect a positive peak at that time, reflecting the polarity inversion of the N_1 component. This was not the case in the present study. Present observations parallel previous studies of AEPs recorded from PLST (e.g., Howard et al., 2000; Steinschneider et al., 2011), wherein the sequence of the most prominent peaks was an initial positivity (P_α), followed by an initial negativity (N_α), and a second positivity (P_β). Furthermore, when the AEPs from all subjects were averaged together, the waveform representing the grand average response on PLST showed the P_α-N_α-P_β-N_β peak sequence, as reported by Howard and colleagues (2000) (see Fig. 4).

The precise etiologies of the large inter-subject variability are unclear, but may represent in part the complexities of gross anatomy in this region (Zilles et al., 1997; Leonard et al., 1998; Hackett, 2007) that would result in variable contributions of radial and tangential sources to the AEPs. By clinical necessity, electrode coverage was non-uniform across subjects, possibly contributing to the observed variability. This variability would be increased if there was uneven coverage over the various functionally distinct subdivisions of non-primary auditory cortex on PLST. Finally, it is possible that variability in the processing

of sounds contributed to variations in AEP amplitude and morphology (Boatman and Miglioretti, 2005).

The time course of high gamma activity was much less variable than that of the AEP across subjects (see Fig. 4B). On average, peak activity occurred on the falling phase of the N_{α} peak of the AEP, consistent with earlier findings (Steinschneider et al., 2011). Sources of variability likely include differences in sensitivity of different cortical sites to specific acoustic attributes of speech sounds and tones, state of arousal of the subject during the passive-listening paradigm, differences in electrode coverage and response timings within PLST (Nourski et al., 2014a). In the latter study, high gamma activity was shown to have the shortest onset latency in the middle portion of PLST, increasing in both anterior and posterior directions along the gyrus. Overall, the current data support the conclusion that both the AEP and high gamma activity primarily represent local cortical activity on PLST, though contributions for volume-conducted activity from superior temporal plane, especially for the AEP, cannot be ruled out.

Overall, findings from intracranial AEP data parallel those observed in the scalp-recorded AEP. It has been suggested that PLST is the main generator of the T-complex as recorded with temporal scalp electrodes (Wood and Wolpaw; 1982; Scherg et al., 1989; Steinschneider et al., 2011). The amplitude of both the scalp-recorded T-complex and the AEP recorded directly from PLST is greater overlying the right hemisphere (Wolpaw and Penry, 1975; see Fig. 5). The amplitude of the T-complex was greater over the right hemisphere when compared to the left even with ipsilateral stimulation. Further, the marked variability of the intracranially-recorded AEP morphology across subjects (see Fig. 4) parallels that seen in the scalp-recorded T-complex (Wolpaw and Penry, 1975). In turn, this variability may limit the utility of assessment of T-complex amplitude and morphology at the single-subject level. However, at the group level, the T-complex can reflect normal development of the auditory cortex (Ponton et al., 2002; Mahajan and McArthur, 2013), cortical plasticity (Itoh et al., 2012), phonological processing (Wagner et al., 2013) and has been shown to be aberrant in language disorders such as specific language impairment (Shafer et al., 2011; Bishop et al., 2012) and dyslexia (Hämäläinen et al., 2011). Therefore, assessment of the AEP recorded directly from PLST offers the opportunity to clarify the underlying detailed neural events and their spatiotemporal organization reflected in the T-complex.

Stimulus classification based on LFP and high gamma ERBP

As a rule, classification accuracy using the mean of either LFP, high gamma ERBP or both peaked at 100–200 ms. This timing for discrimination of stimulus attributes parallels multiple previous findings. Using LFPs, Chang et al. (2010) showed the most robust discrimination of stop consonants at around 110 ms. Similarly, Nourski et al. (2014c) showed peaks in classification accuracy for pure tones at around 100 ms using the distribution of high gamma activity across perisylvian cortex. The importance of this time frame extends to non-invasive studies, including those showing that accurate perceptual encoding of speech in complex listening environments peaks at around 100 ms (e.g. Ding and Simon, 2012).

Classification using the lower-frequency components embedded in the LFP was superior to that based on high gamma ERBP. This observation held across all classification tasks (voicing, POA, tone). Further, addition of information contained in high gamma activity only minimally improved classification accuracy compared to that based solely on the LFP. These results are surprising, given that spatial patterns of high gamma activity within PLST have been shown to reflect phonemic identity (Mesgarani et al., 2014) and accurately discriminate pure tone stimuli (Nourski et al., 2014c). However, these intracranial studies did not concurrently compare the information provided by lower-frequency components embedded in the LFP. Therefore, it is unclear whether high gamma activity would have been superior to the low-frequency LFP if both analyses were conducted. Importantly, the study of Chang et al. (2010) did focus analysis on LFPs and was able to demonstrate categorical differences in the representation of stop consonants based on POA. By extension, non-invasive event-related potentials and neuromagnetic responses examining lower-frequency components of neural activity have been capable of precise classification of speech and other complex sounds (e.g. Alain et al., 2002; Luo and Poeppel, 2007; Toscano et al., 2010; Ding and Simon, 2012).

In the current study, classification of voicing and classification of tones were superior to that based on POA. More natural (and acoustically diverse) stimuli were not examined in our study. We predict that better accuracy would have been obtained when multiple cues normally associated with POA [such as temporally dynamic spectra and voicing onset (e.g. Alexander and Kluender, 2008)], are factored into the analysis. It cannot be excluded, however, that use of more sophisticated classification algorithms or techniques for pre-processing the input may enhance classification accuracy to those identified in previous studies. However, this should be understood in light of the fact that the goal of the SVM-based analysis in the present study was to demonstrate that there is available information in the neural activity, without making assumptions as to how the brain actually identifies and classifies sound stimuli.

Finer-grain characterization of LFP waveform morphology, as captured by the addition of slope, and, to a lesser extent, quadratic, revealed enhanced classification accuracy. This enhancement means that the phase of the LFP, as well as its amplitude, provide relevant information to the classifier. This is consistent with previous observations that the phase of relatively low frequency components of cortical LFPs carries complementary information to that present in amplitude (Luo & Poeppel, 2007; Herrmann et al., 2013; Ng et al., 2013). This was not the case in our characterization of fine-grained time course of high gamma ERBP. This observation likely reflects the relative similarity of high gamma temporal envelopes measured at different sites across voicing and POA contrasts (see Figs. 2B, 3B). Supporting evidence for the importance of slope in the LFP may be observed using current source density analysis performed on LFPs recorded from the primary auditory cortex. This analysis has shown variation in the slope of the rising and falling phase of current sinks based upon the attributes of specific stimuli (e.g., Steinschneider et al., 1998; Happel et al., 2010). By extension, the relevance of the slope of LFPs in the current study might in part reflect the strength of response elicited by any given stimulus at a given cortical site.

When classification analysis was based on high gamma averaged within time windows longer than 50 ms (up to 200 ms), classification accuracy of voicing deteriorated. Adding more detailed temporal descriptors (slope and quadratic) yielded improvement in accuracy at these longer time windows. It is likely that more fine-grained temporal descriptors better captured information about stimulus onset and the later onset of voicing over a more gradual unfolding signal (high gamma activity). The disparity of results between voicing and POA can be explained by the fact that POA information is primarily encoded by spatially distributed patterns of activity elicited by the spectral differences present at syllable onset. While longer time windows might encode formant transition information and allow for a more accurate estimate of these distributed patterns, we used synthetic syllables where formant transition information overlapped across the syllables varying in POA (See Fig. 1). Analyzing classification performance at longer time windows using natural speech might still be beneficial, but this test was beyond the scope of the current manuscript and will have to be examined in subsequent studies.

It is possible that superior classification analysis by LFPs relative to high gamma ERBP might have been due to electrode placement that did not optimally capture sources of prominent high gamma activity in a given subject. To examine this question, we analyzed classification accuracy over time using the best combination of features (LFP and high gamma ERBP) in each subject individually. In all cases, LFP-based classification was superior to high gamma-based analysis and the combination of the two did not enhance classification accuracy. This consistent observation argues against bias of electrode placement that preferentially optimized the capture of the LFP signals versus that seen in high gamma power.

Functional interpretation of LFP and high gamma ERBP

At the cortical level, LFPs primarily represent synaptic rather than neuronal spiking activity. In contrast, high gamma power is a surrogate measure for spiking activity of neuronal aggregates (Friedman-Hill et al. 2000; Frien et al., 2000; Brosch et al., 2002). Thus, the LFP and high gamma power measurements are reflections of different aspects of neural processing within cortex, and may play different roles in representing sound information. In the current study, classification performance was superior when features of the LFP were analyzed. In contrast, in an earlier report, high gamma activity within PLST, but not the AEP, was modulated strongly by attention and behavioral task (Nourski et al., 2014b; Steinschneider et al, 2014).

It must be borne in mind that electrode grids are positioned above lamina 1 in the subdural space. Intracranial studies in experimental animals using current source density analysis have shown that superficially recorded LFPs are dominated by synaptic activity in supragranular layers (Javitt et al., 1994; Steinschneider et al., 1994). These analyses suggest that negativities as recorded from subdural electrodes primarily reflect excitatory postsynaptic potentials within supragranular layers with passive current return occurring primarily at more proximal dendritic sites (Steinschneider et al., 1992). This interpretation suggests that these negativities would be associated with increases in high gamma activity as neurons become depolarized within their apical dendrites. This "up-state" is well-established

in the literature (e.g. Schroeder and Lakatos, 2009; Lakatos et al., 2013), and we suggest that the peak of high gamma activity occurring at or shortly after the N_α peak of the AEP reflects this dynamic (see Fig. 5A).

In conclusion, the current findings bring us back full circle to the relevance of the AEP and low frequency components of the LFP. In the past, analysis of the AEP was preeminent when studying human electrophysiology. Subsequently, a paradigm shift occurred where the use of high frequency gamma activity became preeminent at the expense of the analysis of the AEP. Results suggest that future studies examining intracranial recordings should consider utilizing data obtained from both electrophysiologic features. Both time- and phase-locked low-frequency components as well as high gamma activity provide valuable information regarding sound processing. Further, phase coupling between different cortical regions will likely provide additional information about the ever-changing neural ensembles engaged in speech perception. While the current study only examined the relatively early stages of processing, it is likely that multiple methodologies will have to be utilized to identify the neural stages engaged in higher-order language functions such as syntactic, lexical and semantic processing. It is hoped that these detailed intracranial studies will be of great translational relevance when trying to understand both normal and aberrant electrophysiologic responses obtained in investigations using non-invasive techniques.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Alain C, Schuler BM, McDonald KL. Neural activity associated with distinguishing concurrent auditory objects. J Acoust Soc Am. 2002; 111:990–5. [PubMed: 11863201]

Alexander JM, Kluender KR. Spectral tilt change in stop consonant perception. J Acoust Soc Am. 2008; 123:386–96. [PubMed: 18177168]

Blumstein SE, Stevens KN. Perceptual invariance and onset spectra for stop consonants in different vowel environments. J Acoust Soc Am. 1980; 67:648–62. [PubMed: 7358906]

Boatman DF, Miglioretti DL. Cortical sites critical for speech discrimination in normal and impaired listeners. J Neurosci. 2005; 25:5475–80. [PubMed: 15944375]

Brosch M, Budinger E, Scheich H. Stimulus-related gamma oscillations in primate auditory cortex. J Neurophysiol. 2002; 87:2715–25. [PubMed: 12037173]

Brugge JF, Volkov IO, Oya H, Kawasaki H, Reale RA, Fenoy A, Steinschneider M, Howard MA III. Functional localization of auditory cortical fields of human: click-train stimulation. Hear Res. 2008; 238:12–24. [PubMed: 18207680]

Celesia GG, Puletti F. Auditory cortical areas of man. Neurology. 1969; 19:211–20. [PubMed: 5812977]

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT. Categorical speech representation in human superior temporal gyrus. Nat Neurosci. 2010; 13:1428–32. [PubMed: 20890293]

Chang CC, Lin CJ. LIBSVM: A library for support vector machines. ACM Trans Intell Syst Technol. 2011; 2:27. 1–27:27.

Crone NE, Boatman D, Gordon B, Hao L. Induced electrocorticographic gamma activity during auditory perception. Clin Neurophysiol. 2001; 112:565–582. [PubMed: 11275528]

Ding N, Simon JZ. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A. 2012; 109:11854–9. [PubMed: 22753470]

Edwards E, Soltani M, Deouell LY, Berger MS, Knight RT. High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. J Neurophysiol. 2005; 94:4269–80. [PubMed: 16093343]

Friedman-Hill S, Maldonado PE, Gray CM. Dynamics of striate cortical activity in the alert macaque: I. Incidence and stimulus dependence of gamma-band neuronal oscillations. Cereb Cortex. 2000; 10:1105–16. [PubMed: 11053231]

Frien A, Eckhorn R, Bauer R, Woelbern T, Gabriel A. Fast oscillations display sharper orientation tuning than slower components of the same recordings in striate cortex of the awake monkey. Eur J Neurosci. 2000; 12:1453–65. [PubMed: 10762373]

Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci. 2003; 6:1216–23. [PubMed: 14583754]

Hackett, TA. Organization and correspondence of the auditory cortex of humans and nonhuman primates. In: Kaas, JH., editor. Evolution of Nervous Systems Vol. 4: Primates. Academic Press; New York: 2007. p. 109-119.

Halgren E, Baudena P, Clarke JM, Heit G, Liégeois C, Chauvel P, Musolino A. Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. Electroencephalogr Clin Neurophysiol. 1995; 94:191–220. [PubMed: 7536154]

Hämäläinen JA, Fosker T, Szücs D, Goswami U. N1, P2 and T-complex of the auditory brain event-related potentials to tones with varying rise times in adults with and without dyslexia. Int J Psychophysiol. 2011; 81:51–9. [PubMed: 21565226]

Happel MF, Jeschke M, Ohl FW. Spectral integration in primary auditory cortex attributable to temporally precise convergence of thalamocortical and intracortical input. J Neurosci. 2010; 30:11114–27. [PubMed: 20720119]

Herrmann B, Henry MJ, Grigutsch M, Obleser J. Oscillatory phase dynamics in neural entrainment underpin illusory percepts of time. J Neurosci. 2013; 33:15799–809. [PubMed: 24089487]

Howard MA, Volkov IO, Mirsky R, Garell PC, Noh MD, Granner M, Damasio H, Steinschneider M, Reale RA, Hind JE, Brugge JF. Auditory cortex on the posterior superior temporal gyrus of human cerebral cortex. J Comp Neurol. 2000; 416:76–92.

Itoh K, Okumiya-Kanke Y, Nakayama Y, Kwee IL, Nakada T. Effects of musical training on the early auditory cortical representation of pitch transitions as indexed by change-N1. Eur J Neurosci. 2011; 36:3580–92. [PubMed: 22958242]

Javitt DC, Steinschneider M, Schroeder CE, Vaughan HG Jr, Arezzo JC. Detection of stimulus deviance within primate primary auditory cortex: intracortical mechanisms of mismatch negativity (MMN) generation. Brain Res. 1994; 667:192–200. [PubMed: 7697356]

Kajikawa Y, Schroeder CE. How local is the local field potential? Neuron. 2011; 72:847–58. [PubMed: 22153379]

Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. J Acoust Soc Am. 1990; 87:820–857. [PubMed: 2137837]

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE. The spectrotemporal filter mechanism of auditory selective attention. Neuron. 2013; 77:750–61. [PubMed: 23439126]

Lee YS, Lueders H, Dinner DS, Lesser RP, Hahn J, Klem G. Recording of auditory evoked potentials in man using chronic subdural electrodes. Brain. 1984; 107:115–31. [PubMed: 6697149]

Leonard CM, Puranik C, Kuldau JM, Lombardino LJ. Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? Cereb Cortex. 1998; 8:397–406. [PubMed: 9722083]

Liégeois-Chauvel C, Musolino A, Chauvel P. Localization of the primary auditory area in man. Brain. 1991; 114:139–151. [PubMed: 1900211]

Liégeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. Electroencephalogr Clin Neurophysiol. 1994; 92:204–14. [PubMed: 7514990]

Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 2007; 54:1001–10. [PubMed: 17582338]

Mahajan Y, McArthur G. Maturation of the auditory t-complex brain response across adolescence. Int J Dev Neurosci. 2013; 31:1–10. [PubMed: 23078912]

Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 2012; 485:233–6. [PubMed: 22522927]

Mesgarani N, David SV, Fritz JB, Shamma SA. Phoneme representation and classification in primary auditory cortex. J Acoust Soc Am. 2008; 123:899–909. [PubMed: 18247893]

Mesgarani N, Cheung C, Johnson K, Chang EF. Phonetic feature encoding in human superior temporal gyrus. Science. 2014; 343:1006–10. [PubMed: 24482117]

Mukamel R, Fried I. Human intracranial recordings and cognitive neuroscience. Annu Rev Psychol. 2012; 63:511–37. [PubMed: 21943170]

Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R. Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. Science. 2005; 309:951–4. [PubMed: 16081741]

Ng BS, Logothetis NK, Kayser C. EEG phase patterns reflect the selectivity of neural firing. Cereb Cortex. 2013; 23:389–98. [PubMed: 22345353]

Nir Y, Fisch L, Mukamel R, Gelbard-Sagiv H, Arieli A, Fried I, Malach R. Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. Curr Biol. 2007; 17:1275–85. [PubMed: 17686438]

Nourski KV, Brugge JF. Representation of temporal sound features in the human auditory cortex. Rev Neurosci. 2011; 22:187–203. [PubMed: 21476940]

Nourski KV, Howard MA III. Invasive recordings in the human auditory cortex. Hand Clin Neurol. 2015; 129:225–244.

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA 3rd, Brugge JF. Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci. 2009; 29:15564–74. [PubMed: 20007480]

Nourski KV, Brugge JF, Reale RA, Kovach CK, Oya H, Kawasaki H, Jenison RL, Howard MA III. Coding of repetitive transients by auditory cortex on posterolateral superior temporal gyrus in humans: an intracranial electrophysiology study. J Neurophysiol. 2013; 109:1283–95. [PubMed: 23236002]

Nourski KV, Steinschneider M, McMurray B, Kovach CK, Oya H, Kawasaki H, Howard MA III. Functional organization of human auditory cortex: Investigation of response latencies through direct recordings. Neuroimage. 2014a; 101:598–609. [PubMed: 25019680]

Nourski KV, Steinschneider M, Oya H, Kawasaki H, Howard MA III. Modulation of response patterns in human auditory cortex during a target detection task: An intracranial electrophysiology study. Int J Psychophysiol. 2014b10.1016/j.ijpsycho.2014.03.006

Nourski KV, Steinschneider M, Oya H, Kawasaki H, Jones RD, Howard MA. Spectral organization of the human lateral superior temporal gyrus revealed by intracranial recordings. Cereb Cortex. 2014c; 24:340–52. [PubMed: 23048019]

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF. Reconstructing speech from human auditory cortex. PLoS Biology. 2012; 10:e1001251. [PubMed: 22303281]

Ponton C, Eggermont JJ, Khosla D, Kwong B, Don M. Maturation of human central auditory system activity: separating auditory evoked potentials by dipole source modeling. Clin Neurophysiol. 2002; 113:407–20. [PubMed: 11897541]

Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E. Sequential processing of lexical, grammatical, and phonological information within Broca's area. Science. 2009; 326:445–49. [PubMed: 19833971]

Scherg M, Vajsar J, Picton TW. A source analysis of the late human auditory evoked potentials. J Cogn Neurosci. 1989; 1:336–55. [PubMed: 23971985]

Schroeder CE, Lakatos P. The gamma oscillation: master or slave? Brain Topogr. 2009; 22:24–6. [PubMed: 19205863]

Shafer VL, Schwartz RG, Martin B. Evidence of deficient central speech processing in children with specific language impairment: the T-complex. Clin Neurophysiol. 2011; 122:1137–55. [PubMed: 21147550]

Sinai A, Crone NE, Wied HM, Franaszczuk PJ, Miglioretti D, Boatman-Reich D. Intracranial mapping of auditory perception: event-related responses and electrocortical stimulation. Clin Neurophysiol. 2009; 120:140–9. [PubMed: 19070540]

Steinschneider M, Tenke CE, Schroeder CE, Javitt DC, Simpson GV, Arezzo JC, Vaughan HG Jr. cellular generators of the cortical auditory evoked potential initial component. Electroencephalogr Clin Neurophysiol. 1992; 84:196–200. [PubMed: 1372236]

Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG Jr. Speech-evoked activity in primary auditory cortex: effects of voice onset time. Electroencephalogr Clin Neurophysiol. 1994; 92:30–43. [PubMed: 7508851]

Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA III. Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. J Neurophysiol. 1999; 82:2346–57. [PubMed: 10561410]

Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, Howard MA III. Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. Cereb Cortex. 2005; 15:170–186. [PubMed: 15238437]

Steinschneider M, Fishman YI, Arezzo JC. Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. Cereb Cortex. 2008; 18:610–25. [PubMed: 17586604]

Steinschneider M, Nourski KV, Kawasaki H, Oya H, Brugge JF, Howard MA III. Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus. Cereb Cortex. 2011; 21:2332–2347. [PubMed: 21368087]

Steinschneider M, Nourski KV, Fishman YI. Representation of speech in human auditory cortex: is it special? Hear Res. 2013; 305:57–73. [PubMed: 23792076]

Steinschneider M, Nourski KV, Rhone AE, Kawasaki H, Oya H, Howard MA III. Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. Front Neurosci. 2014; 8:240. [PubMed: 25157216]

Toscano JC, McMurray B, Dennhardt J, Luck S. Continuous perception and graded categorization electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. Psychol Sci. 2010; 21:1532–40. [PubMed: 20935168]

Tsunada J, Lee JH, Cohen YE. Representation of speech categories in the primate auditory cortex. J Neurophysiol. 2011; 105:2634–46. [PubMed: 21346209]

Vaughan HG Jr, Ritter W. The sources of auditory evoked responses recorded from the human scalp. Electroencephalogr Clin Neurophysiol. 1970; 28:360–7. [PubMed: 4191187]

Wagner M, Shafer VL, Martin B, Steinschneider M. The effect of native-language experience on the sensory-obligatory components, the P1-N1-P2 and the T-complex. Brain Res. 2013; 1522:31–7. [PubMed: 23643857]

Wolpaw JR, Penry JK. A temporal component of the auditory evoked response. Electroencephalogr Clin Neurophysiol. 1975; 39:609–20. [PubMed: 53139]

Wood CC, Wolpaw JR. Scalp distribution of human auditory evoked potentials. II. Evidence for overlapping sources and involvement of auditory cortex. Electroencephalogr Clin Neurophysiol. 1982; 54:25–38. [PubMed: 6177515]

Zilles K, Schleicher A, Langemann C, Amunts K, Morosan P, Palomero-Gallagher N, Schormann T, Mohlberg H, Bürgel U, Steinmetz H, Schlaug G, Roland PE. Quantitative analysis of sulci in the human cerebral cortex: development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. Hum Brain Mapp. 1997; 5:218–21. [PubMed: 20408218]

Highlights

- Local field potentials (LFP) provide greater classification accuracy than high gamma.

- Classification accuracy using LFPs is enhanced by adding fine temporal detail.

- Highest accuracy is achieved when using both measures.
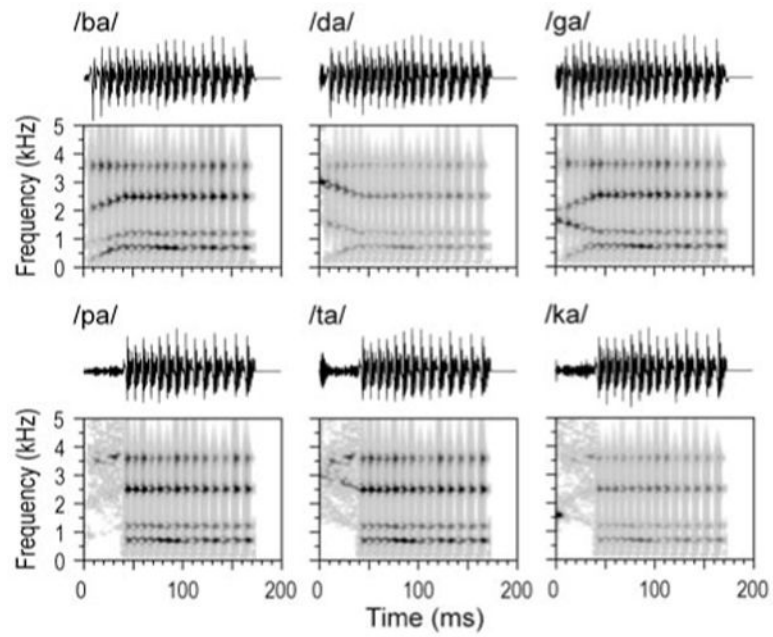
Figure 1.
Waveforms and spectrograms (top and bottom rows, respectively) of the six CV syllables used in the study.
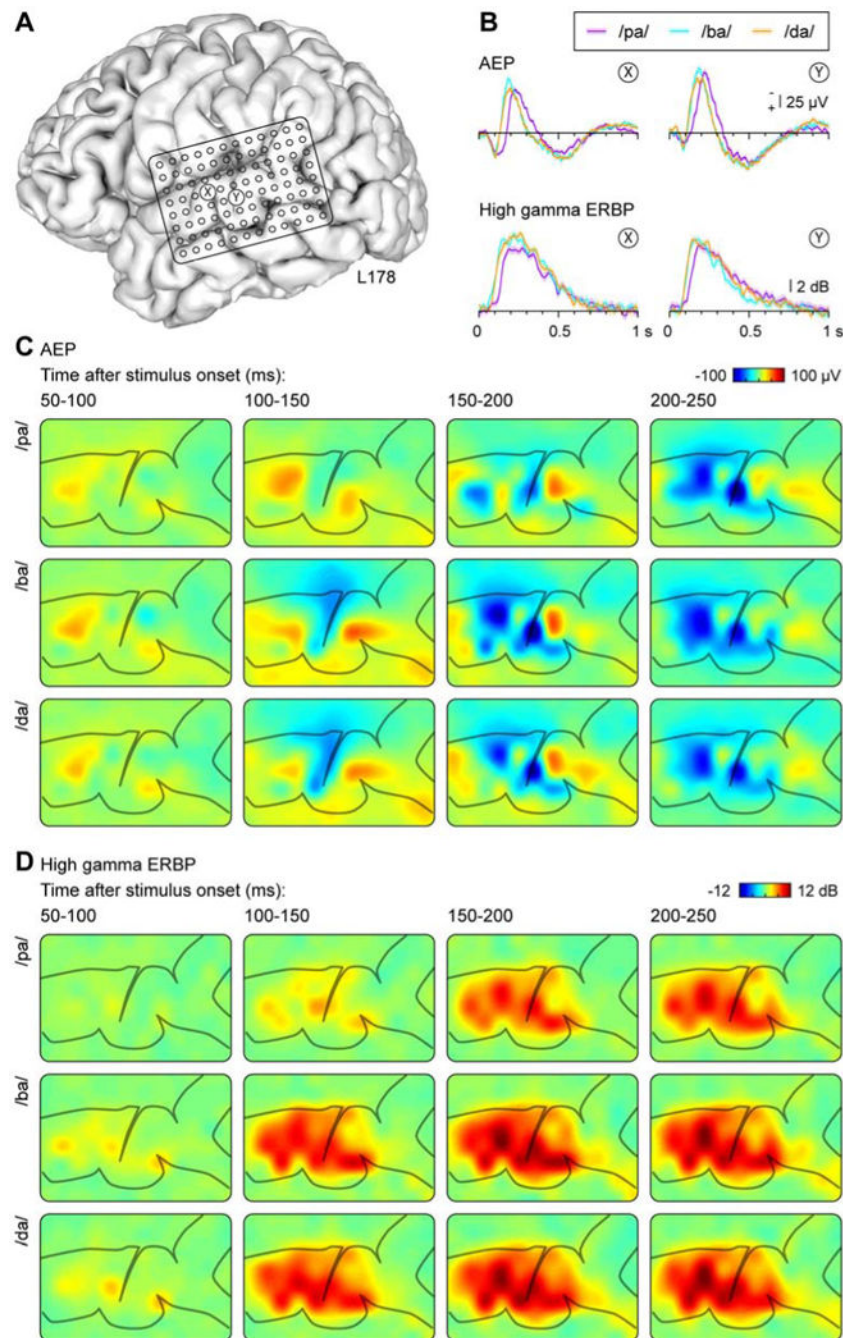
Figure 2.
Responses to CV syllables from the dominant (left) hemisphere in a representative subject (L178). A: Location of the 96-contact subdural grid. B: AEP and high gamma ERBP recorded from two exemplary sites (X and Y, location marked in panel A) in response to syllables /pa/, /ba/, and /da/. C, D: Cortical activation maps showing AEP and high gamma ERBP (C and D, respectively), elicited by the three CV syllables (rows) and averaged within 50 ms consecutive time windows (columns).
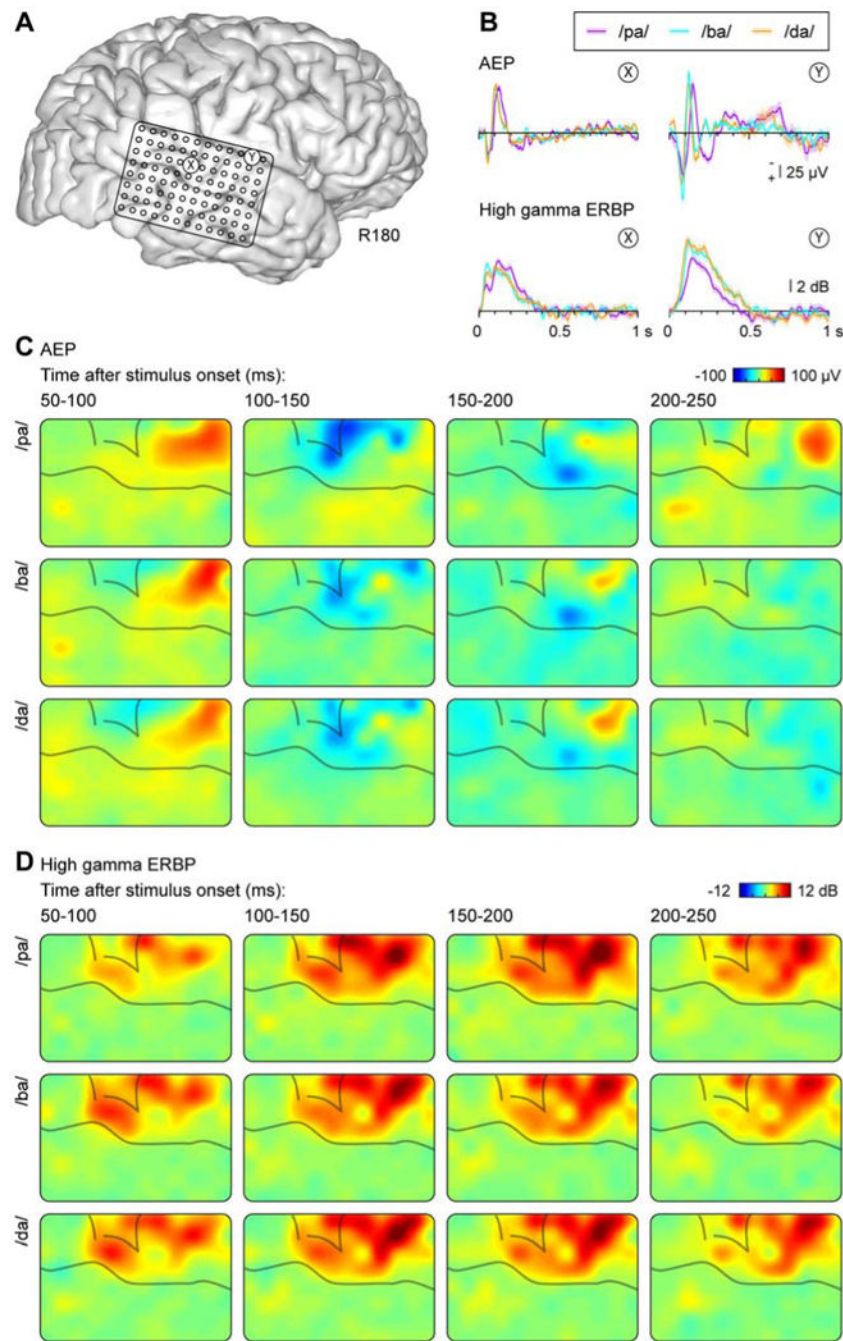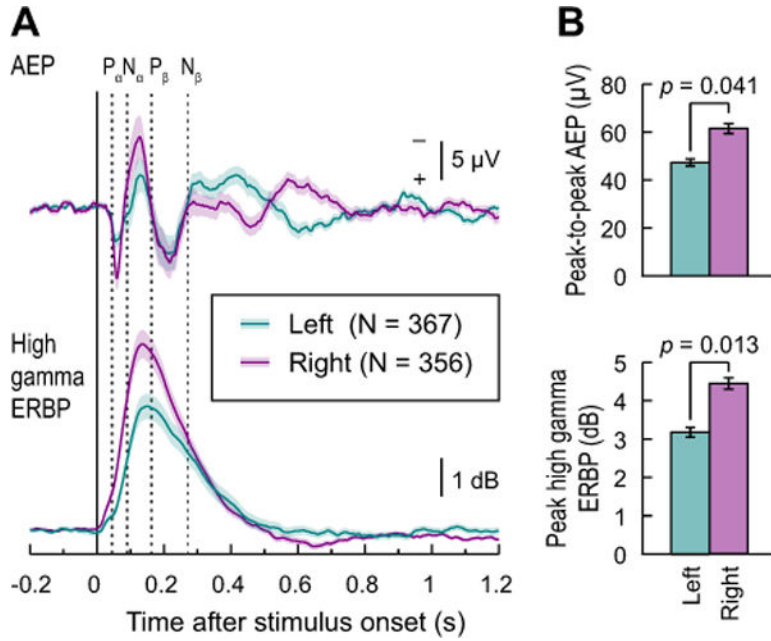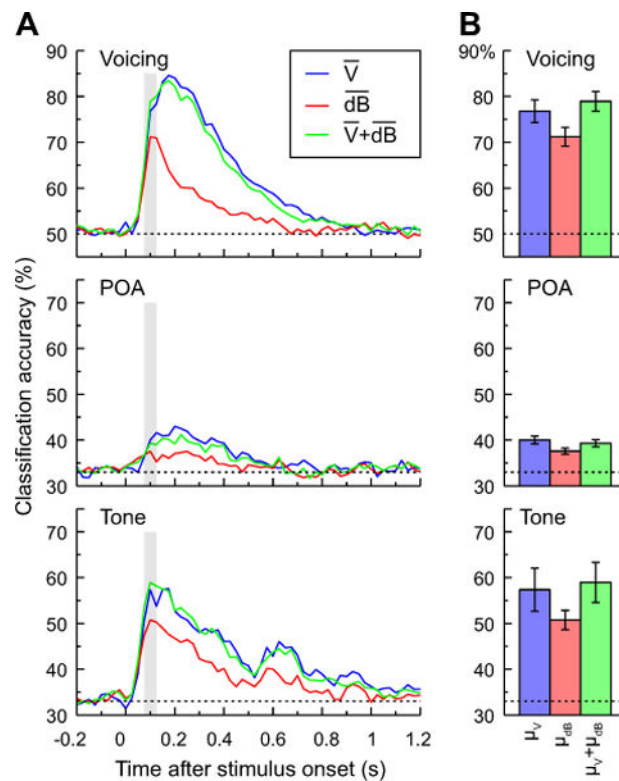
Figure 3.
Responses to CV syllables from the non-dominant (right) hemisphere in a representative
subject (R180). See legend of Figure 2 for details.
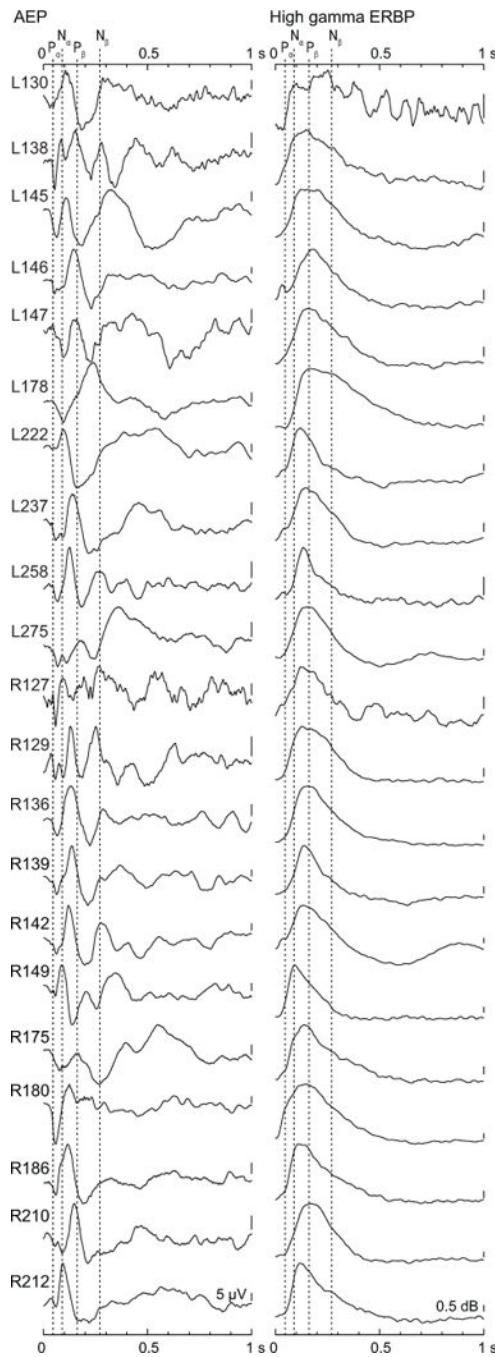
Figure 4.
Morphology of AEP and high gamma ERBP waveforms, averaged across responses to voiced CV syllables (/ba/, /da/, and /ga/) over recording sites overlying STG in each subject (rows). Dashed lines denote mean latencies of P_α, N_α, P_β and N_β components of AEP waveforms elicited by click train stimuli, as described by Howard et al. (2000).
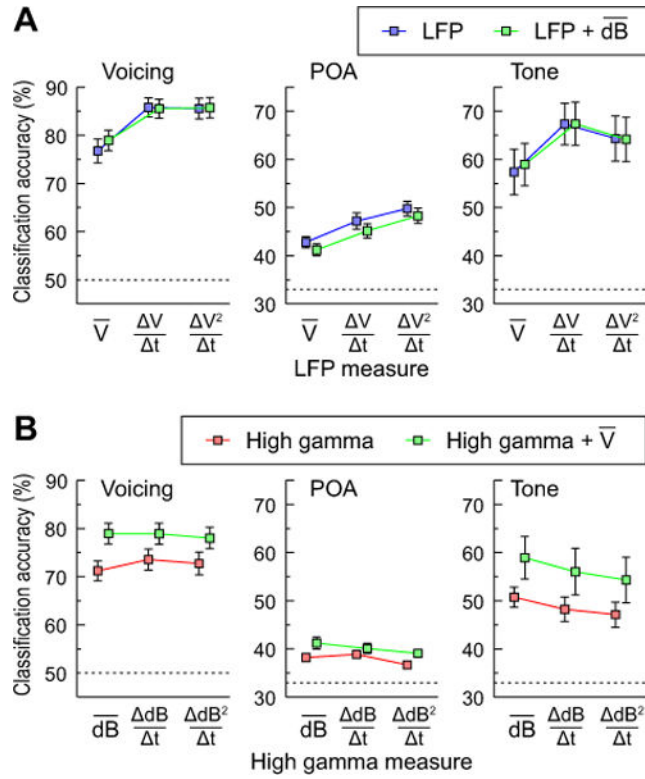
Figure 5.
Morphology and variability of AEP and high gamma ERBP waveforms. A: Grand average AEP (top) and high gamma ERBP (bottom) and their 95% confidence intervals are shown for left and right hemisphere sites (teal and purple, respectively). Plots represent responses to voiced CV syllables (/ba/, /da/, and /ga/), recorded from sites overlying STG in all subjects. Dashed lines denote mean latencies of P_α, N_α, P_β and N_β components of AEP waveforms elicited by click train stimuli, as described by Howard et al. (2000). B: comparison of peak-to-peak AEP amplitudes (top) and peak high gamma ERBP (bottom). Mean values and S.E.M. are plotted for left and right hemisphere sites in teal and purple, respectively. Significance was established using independent-samples *t*-tests.
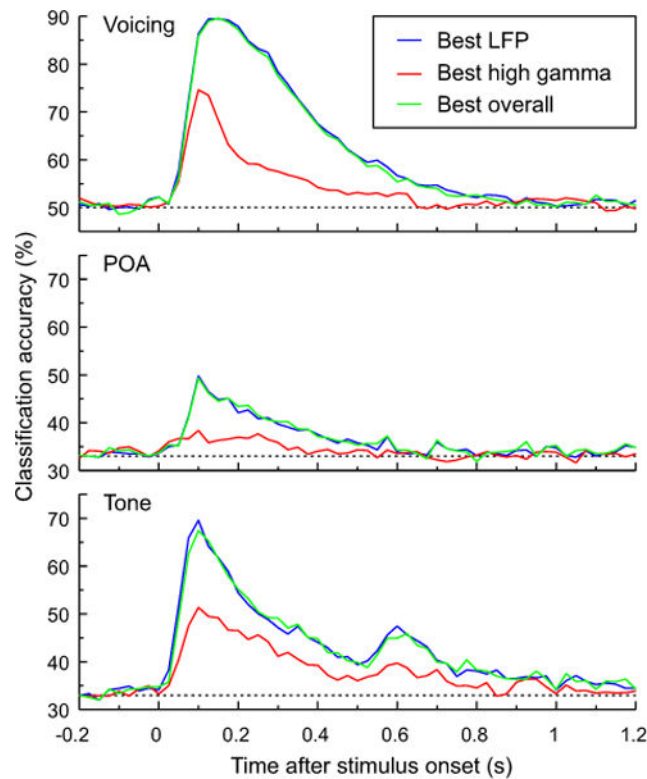
Figure 6.
Classification analysis of CV syllables based on voicing and POA (top and middle rows) and pure tones (bottom row), based on LFP, high gamma ERBP, or both measures, averaged within 50 ms windows (plotted in blue, red and green, respectively). A: Classification accuracy as a function of time window averaged across all subjects ($N = 21$ for voicing and POA, $N = 10$ for tone classification). Vertical shaded bars denote the 75–125 ms window used for subsequent analyses. B: Mean classification accuracy for the three tasks within the 75–125 ms window and across-subject S.E.M. Dashed lines indicate chance level (50% for voicing and 33% for POA and tone classification).

Figure 7.
Contribution of LFP and high gamma measures to classification accuracy in the 75–125 ms window; classification of CV syllables based on voicing and POA (left and middle column) and pure tones (right column). A: Effects of adding slope and quadratic to mean LFP with or without adding mean high gamma ERBP. B: Effects of adding slope and quadratic to mean high gamma ERBP with or without adding mean LFP. Error bars denote across-subject S.E.M. Dashed lines indicate chance level (50% for voicing and 33% for POA and tone classification).

Figure 8.
Classification analysis of CV syllables based on voicing and POA (top and middle panels) and pure tones (bottom panel). Across-subject ($N = 21$ for voicing and POA, $N = 10$ for tone classification) average accuracy based on LFP and/or high gamma measures that yielded the best performance in each subject. Dashed lines indicate chance level (50% for voicing and 33% for POA and tone classification).

Table 1

One sample t-tests comparing classification accuracy at 75–125 ms after stimulus onset to chance.

Measure	Voicing		POA		Tone	
	$T(20)$	p	$T(20)$	p	$T(9)$	p
Mean LFP $(\overline{V})$	10.83	<0.0001	8.08	<0.0001	5.07	0.007
Mean high gamma $(\overline{dB})$	10.92	<0.0001	5.06	<0.0001	7.82	<0.0001
$(\overline{V})+(\overline{dB})$	13.92	<0.0001	7.14	<0.0001	6.10	0.0002