Single-Trial Speech Suppression of Auditory Cortex Activity in Humans

A. Flinker[1], **E.F. Chang**[3], **H.E. Kirsch**[4], **N.M. Barbaro**[3], **N.E. Crone**[5], and **R.T. Knight**[1,2,3,4]

[1]Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA

[2]Department of Psychology, University of California, Berkeley, CA, USA

[3]Department of Neurosurgery, University of California, San Francisco, CA, USA

[4]Department of Neurology, University of California, San Francisco, CA, USA

[5]Department of Neurology, The Johns Hopkins University, Baltimore, MD, USA

Abstract

The human auditory cortex is engaged in monitoring the speech of interlocutors as well as self-generated speech. During vocalization, auditory cortex activity is reported to be suppressed, an effect often attributed to the influence of an efference copy from motor cortex. Single unit studies in non-human primates have demonstrated a rich dynamic range of single-trial auditory responses to self-speech consisting of suppressed, non-suppressed and excited auditory neurons. However, human research using non-invasive methods has only reported suppression of averaged auditory cortex responses to self-generated speech. We addressed this discrepancy by recording electrocorticographic activity from neurosurgical subjects performing auditory repetition tasks. We observed that the degree of suppression varied across different regions of auditory cortex, revealing a variety of suppressed and non-suppressed responses during vocalization. Importantly, single-trial high gamma power (γ_{High}: 70-150 Hz) robustly tracked individual auditory events and exhibited stable responses across trials for suppressed and non-suppressed regions.

Keywords

electrocorticography; auditory processing; self-monitoring; speech production; auditory feedback

Introduction

During speech production we continuously monitor our own voice and compensate for changes in auditory feedback (Levelt, 1983). For example, speakers change their voice both in intensity and pitch when they are introduced to a noisy acoustic environment (Lane and Tranel, 1971). Furthermore, delaying a speaker's auditory feedback will disrupt fluent speech production (Yates, 1963). Despite the importance of auditory feedback for accurate production, it remains unclear how auditory cortex processes self-generated speech during vocalization.

Single-unit studies in non-human primates as well as humans have reported suppressed auditory neuronal responses during vocalization (Müller-Preuss and Ploog, 1981; Creutzfeldt et al., 1989). Although many of the single-unit responses showed a marked

suppression in activity, a large population of auditory neurons exhibited an excited response to self-generated vocalization. Recent work with vocalizing marmosets has reported auditory neurons with a varying degree of suppressed responses. A majority of neurons showed some type of suppression while a smaller number exhibited excited responses. These results suggested that while auditory neurons showed a spectrum of responses, the average of the population exhibited a suppressed response (Eliades and Wang, 2005; Eliades and Wang, 2008).

Non-invasive investigations of human auditory responses during vocalization have only reported averaged suppressed responses using functional imaging and electrophysiological studies (Numminen et al., 1999; Wise et al., 1999; Ford et al., 2001; Houde et al., 2002; Christoffels et al., 2007). Electrophysiological studies have reported suppression in the N100 and M100 components of auditory-evoked potentials peaking at 100 ms (Numminen et al., 1999; Ford et al., 2001).

The suppression of auditory cortex during vocalization has often been attributed to the influence of motor cortex. Current theories present a forward model where corollary discharge signals, representing a prediction of impending self-generated stimuli, modulate auditory cortex activity (Ford et al., 2001; Houde et al., 2002). Recent single-unit work has shown that normally suppressed auditory neurons enhanced their activity when auditory feedback was altered (Eliades and Wang, 2008), in accord with evidence from human EEG and MEG studies (Houde et al., 2002; Heinks-Maldonado et al., 2005). Although speech suppression in human auditory cortex is well accepted, the temporal dynamics of suppression and its stability at the level of single-trials remains unknown. Furthermore, the spatial distribution and variability of auditory cortex suppression are unknown.

Electrocorticographic (ECoG) signals acquired directly from the surface of the human cortex have a high signal-to-noise ratio ideal for single-trial analysis and provide a better spatial sampling of neuronal populations than scalp EEG. ECoG studies to date have only shown an averaged suppression of γ_{High} band (γ_{High} >70Hz) power responses (Crone et al., 2001; Towle et al., 2008). Although γ_{High} has been linked to single unit and BOLD activity (Mukamel et al., 2005; Allen et al., 2007; Belitski et al., 2008; Sohal et al., 2009), it is unclear whether the suppression occurs in other frequency bands and how it changes over trials. Similarly, it is not known if suppression is uniform across auditory cortex or instead shows a regional topography of the degree of suppression.

Results

Auditory Responses Across Subjects

In order to assess auditory cortex responses to speech during listening and production we first selected electrodes with clear auditory spectral responses. Electrodes were selected if they covered the superior temporal gyrus (STG) and showed a statistically significant auditory spectral response in any frequency band (1-300 Hz, including power of the raw signal) compared with baseline (p<0.001, t-test Bonferroni corrected). Event-related potentials (ERPs) for both listening to vowels and producing the same vowels were computed for each auditory electrode and averaged per subject. A grand average across all subjects is shown in Figure 1(a). A negative potential peaking ~150 ms is evident for the hearing ERP (solid trace) and this response is severely reduced for the speaking ERP (dashed trace; t-test, p< 0.05). This finding is similar to previous scalp EEG studies reporting a reduction in the auditory N100 ERP component (Ford et al., 2001; Heinks-Maldonado et al., 2005).

We assessed auditory responses in the frequency domain by comparing auditory spectral responses during hearing vowels with spectral responses during production of the same vowels. In each subject four to eight auditory electrodes exhibiting a significant spectral response were selected for analysis (same selection criteria described above). Power in different frequency bands, averaged during a 300 ms post-stimulus-onset time window, was measured while hearing vs. speaking vowels and a difference signal was computed for each subject. This signal was computed by taking the average difference in power between the hearing and speaking window after normalization to a pre-stimulus baseline. Figure 1(b) shows the difference signals averaged across subjects, representing the degree of suppression in each frequency band. Maximal suppression was found in the 80 Hz, 100 Hz and 120 Hz bands (70-130 Hz: t-test, $p<0.01$ for 70-90 Hz band and $p<0.001$ for 90-110, 110-130 Hz bands). In the lower frequencies, power in the theta band (4-7 Hz) was also suppressed, though with higher variability across subjects (t-test, $p<0.05$). The raw power (0.1-300 Hz) exhibited lower suppression values, which nevertheless passed significance threshold (t-test, $p<0.05$).

Suppression onset and peak

In order to examine the temporal dynamics of the suppressed responses during speech we computed a power time series for each electrode and frequency band using a sliding window approach (50 ms window, 50% overlap). Suppression onset and peak were calculated for each time series exhibiting significantly larger activity during hearing compared with speaking for at least 3 consecutive time-points, equivalent to 100 ms ($p<0.05$, t-test FDR corrected). Only frequency bands within the γ_{High} range showed consistent suppression across all the subjects with a mean onset time of 89.5 ms and a mean peak time of 173 ms across the γ_{High} frequency bands (Figure S1). The 90-110 Hz band had the largest number of electrodes exhibiting suppression with a mean onset time of 104 ms and a mean peak time of 176 ms. The suppression onset times in this band were not significantly different across electrode anatomical location ($F(2,21)=1.3$ $p=0.29$, one way analysis of variance). Lower frequency bands exhibited less consistent results across subjects with two bands exhibiting a small number of suppression onset times prior to articulation (Theta and Alpha bands, see Figure S1).

Response Variability Across STG

We assessed the spatial distribution of auditory responses across the STG focusing on the γ_{High} band (γ_{High}: 70-150 Hz), which showed the maximal degree of suppression. We found a wide distribution of the degree of suppression across the STG. Figure 2 depicts subject S3 with typical auditory responses to hearing (red traces) and speaking (blue traces) vowels. While three electrodes show a robust suppression during speaking compared with hearing the same vowels, two adjacent electrodes exhibit only mild to no suppression. We quantified the degree of suppression for all γ_{High} auditory electrodes using a suppression index (SI) varying from 1 (completely suppressed) to -1 (completely enhanced). Figure 3 shows a wide spectrum of responses with a varying degree of suppression in the different auditory electrodes. The spatial distribution of the SI for each subject is shown in Figure 4. Each individual subject exhibited a regional topography of suppressed auditory responses, which varied spatially across the STG and remained stable across trials.

In order to examine the spatial distribution of suppression every electrode was classified as posterior, anterior or central (within 1 cm) to the lateral surface of Heschl Gyrus. The suppression values across the three spatial groups did not differ statistically ($F(2,33)=0.26$ $p=0.77$, one way analysis of variance). Interestingly, the two electrodes which showed evidence of excited responses during self-generated speech were posterior to Heschl Gyrus (see Figure S2). Three of the seven subjects performed a visual reading task in order to rule

out repetition suppression effects. γ_{High} activity during production of auditory stimuli did not differ significantly from γ_{High} activity during production of visual stimuli (Wilcoxon rank-sum, $p>0.05$; see Figure S3).

Single-Trial Responses

We investigated single-trial auditory responses in all subjects. Figure 4 shows stacked single power traces from a representative electrode for each subject. The single-trial traces track the induced spectral responses to the spoken auditory stimuli, which on average commenced at 100 ms and lasted for 300 ms. Single-trial responses locked to hearing vowels are robust across trials while responses during production (marked in black) are consistently suppressed. This demonstrates a robust and stable response pattern during both the hearing and production phase of the task at the single-trial level. In order to quantify the auditory response fidelity of the single-trials we ran a bootstrapping procedure to compare single-trial auditory responses during hearing and speaking with the pre-stimulus baseline. Figure 5(a) shows responses to hearing (red) and speaking (blue) for all γ_{High} auditory electrodes in units of z-scores compared to the bootstrapped baseline power distribution. A majority of hearing events (red) crossed significance threshold (65.3% of events across all electrodes were significant, $p<0.05$). Speaking events (blue) were significant for some electrodes but on average were less robust (27.1 % of events across all electrodes were significant, $p<0.05$). Further examination of all the individual single-trial events revealed a linear relationship ($R = 0.55$, $p<0.001$) between the magnitude of γ_{High} responses during hearing and speaking as shown in Figure 5(b). This relationship held within all individual subjects (Figure S3).

Independent Spatial Responses

The anatomical distribution of responses across the STG provides evidence for independent signals at the 1-cm resolution of most of our grids. We report data from one subject implanted with a high-density grid with 4 mm inter-electrode spacing over STG (Figure 6(a)). Two adjacent electrodes separated by 4 mm of cortex exhibit functionally distinct responses. Electrode A responds robustly to external stimuli and is suppressed during vocalization (suppression index of 0.36). Electrode B is minimally responsive to external stimuli and is highly selective to self generated speech (suppression index of - 0.3). These functionally distinct patterns of cortical activity at 4-mm separation are also observable at the level of single-trial responses (Figure 6(b)).

Materials and Methods

Subjects

Seven subjects (S1-7) undergoing neurosurgical treatment for refractory epilepsy participated in the study. During clinical treatment the subjects were implanted with one or more electrode arrays with an inter-electrode spacing of 1 cm. Electrode placement and medical treatment were dictated solely by the clinical needs of the patient. Electrophysiological signals were subsequently monitored by clinicians for a period of approximately one week. During lulls in clinical treatment, subjects willing to participate in the study provided written and oral consent. Four male subjects (S1-4, ages 18, 38, 12 and 18 respectively) participated at Johns Hopkins Hospital. One male subject (S5 age 34) and two female subjects (S6-7 ages 33, 51 respectively) participated at UCSF Hospital (see Supplemental Table 1 for pathology details). All subjects were fluent in English as a native tongue and had no language production deficits. Subjects were not receiving anti-epileptic medications during the recording period and were seizure free for at least three hours prior to performing the task. All subjects gave written consent to participate in the study as well as an additional oral consent immediately prior to recording the task. The study protocol was

approved by the UC San Francisco, UC Berkeley and Johns Hopkins Committees on Human Research.

One male subject (S8) participated in a separate pilot study intraoperatively at UCSF while undergoing neurosurgical treatment for tumor resection. The procedure involves one surgical procedure including intraoperative awake language and motor mapping followed by tailored resection of the seizure focus under ECoG guidance. After all clinical mapping was performed, the surgeon placed a high density electrode array with inter-electrode spacing of 4 mm. The subject performed a phoneme repetition task for several minutes after which time the grid was removed and the surgeon continued clinical treatment. The subject provided written and oral consent prior to the surgery and was informed that the task was for research purposes. During surgery, the subject was informed by the surgeon when the clinical mapping was over and the research task was completed under the discretion of the surgeon. The study protocol was approved by the UC San Francisco and UC Berkeley Committees on Human Research.

Task and Stimuli

Seven subjects (S1-7) performed a phoneme repetition task consisting of nine English vowels (/i/, /u/, /I/, /ə/, /o/, /e/, /ʌ/, /æ/, /ɒ/). The stimuli were digitally recorded from a female native speaker of English, acquired at a sampling rate of 44 KHz and 16-bit precision. Recorded stimuli varied in length (215-350 ms) with a mean of 282 ms and standard deviation of 46 ms. The subjects were presented with the digital audio recordings of the vowels via two speakers in front of them. Subjects were instructed that they were going to hear several speech sounds and they were to repeat aloud each speech sound they heard as best they could. The subjects' responses were recorded by up to three different microphones: one close to the mouth, one close to the ear and a third in the ceiling which is part of the clinical video recording system. One microphone closest to the subject was fed directly to the recording system in order to record responses simultaneously with the electrophysiological signals. Similarly, the presented acoustic stimuli signal was sent to the recording system to ensure simultaneous acquisition. The experiment consisted of a total of 72 vowels presented in a pseudorandom fashion with a jittered inter stimulus interval (ISI) of 4 seconds ±250 ms (random jitter). One subject (S8) was part of a pilot study performing a separate task intraoperatively. The task was similar in design and presentation although the stimuli consisted of synthesized /ba/ and /pa/ phonemes.

Subjects S1, S2 and S4 performed an auditory word repetition and visual word reading task in addition to the phoneme repetition task (see Supplemental Material). Auditory and word stimuli consisted of mono- or disyllabic words which were presented via speakers or computer monitor in front of the subject. The subject was instructed to repeat each word they heard during the auditory repetition task and to read aloud each word on screen during the visual reading task.

Electrode Localization

A structural preoperative MRI was acquired for all subjects as well as a post-implantation CT. The MR and CT were reoriented and resliced to a conformed 1 mm space. Using OsiriX Imaging Software, a neurosurgeon marked several anatomical fiducials that were visible both on the CT and MRI (example fiducials include the cerebellar pontene axis, naison, optic nerves, etc.). Once the anatomical markers were placed an affine point based registration was performed in order to localize the CT and MRI in the same space. The fused images were rendered in 3D and assessed for anatomical accuracy (bone structure, visible soft tissue etc.). The 3D render was then compared to an intraoperative image of the exposed grid after it was sutured to the dura. Electrodes covering the superior temporal gyrus and

sylvian fissure were marked according to the 3D reconstruction and the intraoperative image. In one case the subject did not have a CT scan due to an additional surgery performed to reduce swelling. In this case the auditory electrodes were marked by a neurosurgeon based on the intraoperative image alone.

Data Acquisition

Electrophysiological and peripheral auditory channels were acquired using a custom built Tucker Davis Technologies recording system (256 channel amplifier and Z-series digital signal processor board) at the UCSF site and a clinical 128-channel Harmonie system (Stellate, Montreal, Canada) recording system at Johns Hopkins. EEG channels were sampled at 3052 Hz (UCSF) and 1000 Hz (Johns Hopkins) while the peripheral auditory channels were sampled at 24.4 KHz (UCSF) and 1000 Hz (Johns Hopkins). Additional microphones in the room sampled speech at 44 KHz. Electrophysiological data was recorded using a subdural electrode as reference and a scalp electrode as ground. The reference electrode was assigned postoperatively according to clinical needs.

Electrode Selection

Subjects were implanted with 64-100 electrodes covering extensive perisylvian regions varying per subject. For each subject a subset of 8-16 STG (superior temporal gyrus) electrodes were selected based on anatomy. The exact criterion was coverage of the middle through posterior STG and sylvan fissure. For each STG electrode auditory spectral responses were computed for seven different frequency bands (Raw Power: 1-300 Hz, Theta: 4-8 Hz, Alpha: 8-12 Hz, Beta: 12-30 Hz, Gamma: 30-70 Hz, High Gamma: 70-150 Hz, Very High Gamma: 150-300 Hz). Spectral responses were computed by calculating the log transformed power across the entire data time-series and then extracting event-related windows. The log transform was used in order to ensure the data is normally distributed and can be assessed using a t-test. Post-stimulus power was defined as the averaged power across a 300 ms window after hearing onset (0-300 ms) and baseline power was defined as the averaged power across a 300 ms window prior to hearing onset ($-350 -> -50$ ms). An STG electrode was defined as an auditory electrode if it exhibited a statistically significant power response in any of the seven frequency bands. Statistical significance was assessed using a single-tailed two-sample t-test comparing baseline power with post-stimulus power. T-tests were applied with a confidence interval of $p<0.001$ without assuming equal variance (Behrens-Fisher problem) and were corrected for multiple comparisons using Bonferroni correction (accounting for the number of electrodes tested and the number of frequency bands). γ_{High} auditory electrodes refer to auditory electrodes that exhibited a statistically significant power response in the γ_{Hig} band (70-150 Hz).

Data Analysis

All ECoG channels were manually inspected by a neurologist in order to identify channels with interictal and ictal epileptiform activity and artifact. Channels contaminated by epileptiform activity, electrical line noise (60 Hz) or abnormal signal were removed from further analysis. All remaining channels were re-referenced to a common averaged reference defined as the mean of all the remaining channels. Epochs in which ictal activity spread to adjacent channels were removed from further analysis. Speaker and microphone channels recorded simultaneously with ECoG activity were manually inspected in order to mark onset of stimulus and the subsequent response. The audio channels were inspected using both the raw time-series as well as a time-frequency representation (spectrogram) to ensure accurate onset estimation. Trials in which the subject did not respond with a phoneme were removed from analysis; similarly trials overlapping with ictal activity were discarded.

Event related potentials were computed on the band-passed signal (0.1-20 Hz) and were baseline corrected for a 100 ms window ($-100 \rightarrow 0$ ms). Statistical significance was assessed using a paired t-test comparing negative peaks (minimum amplitude within 100-200 ms window) across subjects for hearing versus speaking conditions.

The spectral suppression signal (Fig 1b) was computed for a given electrode and trial by computing the averaged spectral power in a 300 ms time window relative to hearing (0 – 300 ms post hearing onset), speaking (0 – 300 ms post speaking onset) and baseline ($-350 \rightarrow -50$ ms pre hearing onset). The suppression signal was defined as $100*(P_{hear}-P_{speak})/P_{baseline}$, where P is the averaged spectral power for a given time window and frequency band. Frequency bands included the raw signal (1-300 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-30 Hz) and eight different 20 Hz bands between 30 and 190 Hz (i.e., 30-50, 50-70, ... , 170-190 Hz). The suppression signal was averaged across trials and electrodes for each subject and the mean across subjects is plotted in Fig 1(b). Statistical significance was assessed using a one-sample t-test on spectral suppression values across subjects for a given band. The suppression index was calculated by computing the averaged spectral power relative to hearing and speaking (same 300 ms windows as above). The suppression index was defined as $(P_{hear}-P_{speak})/(P_{hear}+P_{speak})$, where P is the averaged spectral power for a given time window.

Single-trial γ_{Hig} traces were computed by first calculating the spectral power time series (70-150 Hz) for the entire block of data. Event related windows of the time series were extracted and transferred to units of percent change compared with baseline (averaged spectral power within $-200 \rightarrow 0$ ms pre-stimulus). Power traces were either averaged across trials (Figure 2) or shown in single-trials (Figure 4). Single-trial analysis across all trials and electrodes (Figure 5) was based on the spectral power in the γ_{High} range (70-150 Hz) during 300 ms time-windows (0-300 relative to hearing and speaking; $-350 \rightarrow -50$ for baseline). Regression analyses (Figures 5(b) and S4) were performed on log-transformed spectral power values (prior to averaging over the time-windows). A statistical bootstrapping procedure was used to compare hearing and speaking spectral responses to baseline within each γ_{High} electrode (Figure 5(a)). For each spectral power condition (hearing, speaking and baseline) two thousand random pairs of single-trials were pooled and the mean of each pair was computed. These two thousand means form a distribution, which is normally distributed and is comparable across conditions. All hearing and speaking bootstrapped values were transformed to units of baseline by subtracting the mean of the baseline distribution and then dividing by the standard deviation of the baseline distribution (z-score). The bootstrapped statistics were done for each individual electrode separately. High gamma spectral responses locked to hearing words as well as producing auditory and visual words in the supplemental material section were computed for all electrodes identified as γ_{High} auditory electrodes in the phoneme repetition task (see Electrode Selection). γ_{High} spectral responses were averaged across electrodes and the mean values across subjects are depicted in Figure S3. Statistical assessment of spectral values within a task (phoneme and word repetition) was performed using a two sampled t-test while assessment between tasks was performed using a non-parametric rank-sum Wilcoxon test.

In the analysis of variance specified in the *Suppression onset and peak* and *Response Variability Across STG* sections of the results, the anatomical location of each electrode was defined as posterior, anterior or central (within 1 cm) to the lateral surface of Heschl Gyrus. Heschl Gyrus was identified and marked manually within each subject.

Suppression onset and peak analysis

Every pair of electrode and frequency band (see Data Analysis section for frequency ranges) that exhibited a significant auditory response during hearing compared with baseline was

selected for analysis (significance was assessed similarly to the Electrode Selection section). For each pair, a spectral power time series was computed, segmented into event related windows ($-200 \rightarrow 400$ ms relative to hearing or speaking onsets) and then resampled using a sliding window approach (50 ms window with 50% overlap). That is, any time point in the new event related time series represents a mean of 50 ms of data and shares 25 ms of data with any neighboring time point. For every electrode and frequency pair there now exists a set (defined by the number of trials) of the resampled hearing event-related time series and speaking event-related time series. Each event-related time point was statistically assessed by comparing a hearing time-point with its speaking counterpart (single-tailed two-sample t-test). Suppression onset was defined as the first of at least 3 consecutive significant time-points. Suppression peak was defined as the maximal value reached after the suppression onset. T-tests were corrected for multiple comparisons (accounting for the number of time points tested) using a False Discovery Rate (FDR) correction of q=0.05 (Benjamini and Hochberg, 1995). Only electrode and frequency pairs with a suppression onset are included in Figure S1.

Spectral Decomposition

Spectral signal analysis was implemented using a frequency domain Gaussian filter (similarly to Canolty et al., 2007). An input signal X was transformed to the frequency domain signal X_f using an N-point fft (where N is defined by the number of points in the time-series X). In the frequency domain a Gaussian filter was constructed (for both the positive and negative frequencies) and multiplied with the signal X_f. The subsequent filtered signal was transformed back to the time-domain using an inverse fft. Power estimates were calculated by taking the Hilbert transform of the frequency filtered signal and squaring the absolute value. All frequency domain filtering and power estimations are comparable to other filtering techniques, such as the wavelet approach (Bruns, 2004).

Discussion

This study addresses the temporal fidelity and spatial topography of auditory cortex suppression during vocalization and resolves a controversy in the animal versus human literature. We first examined averaged ECoG responses across auditory electrodes in seven different subjects. We found a reduction in the N100 component of the ECoG auditory ERP as well as a reduction of induced spectral responses that peaked at 100 Hz, corresponding with the high gamma (γ_{High}) band. However, examining each auditory electrode with a γ_{High} response revealed differential degrees of suppression across auditory cortex. Moreover, within each subject different regions of auditory cortex exhibited different types of self-speech modulation of ECoG auditory responses. Single-trial analysis of these electrodes revealed a consistent response across the different trials. Both highly suppressed and non-suppressed electrodes revealed the same pattern of response in single-trials across the experimental session. Only a few sites demonstrated substantial excitation in the posterior STG during self-generated speech. This finding is in accord with previous single-unit data reported in human and non-human primates (Müller-Preuss and Ploog, 1981; Creutzfeldt et al., 1989; Eliades and Wang, 2003). Lastly, we found a correlation between single-trial responses during speech and hearing suggesting that auditory responses during vocalization, though suppressed, often remain above noise level.

Non-invasive studies in humans have demonstrated suppression of averaged responses from auditory cortex, indexed by a reduction in the N100 and M100 components of auditory ERPs (Numminen et al., 1999; Curio et al., 2000; Ford et al., 2001; Houde et al., 2002). Our finding of a pronounced reduction of a negative ERP component in this latency range during speech supports these non-invasive findings. Auditory ERPs recorded directly from the surface of cortex are much larger in amplitude than scalp ERPs and are likely generated by

local neuronal populations in the STG, although volume conduction from adjacent planum-temporal generators cannot be ruled out.

The signal strength of human EEG spectral power is inversely proportional to the frequency and drops as a function of distance from the cortex (Pritchard, 1992; Freeman, 2004; Bédard et al., 2006). Furthermore, scalp EEG signals are susceptible to both volume conductance effects causing spatial smearing (Nunez P. L., 2005) as well as noise from scalp (Goncharova et al., 2003; Fu et al., 2006), facial (Whitham et al., 2008), and extraocular (Yuval-Greenberg et al., 2008) muscles. In contrast, intracranial recordings largely circumvent these issues providing high SNR, richer spectral content, as well as physical sampling from a well-defined region of cortex.

Two previous electrocorticographic studies have reported a reduction in γ_{High} power during speech (Crone et al., 2001; Towle et al., 2008). These studies used narrow criteria for the γ_{High} band (Towle: 70-100 Hz, Crone: 80-100 Hz) and did not address suppression in other frequency bands. We specifically probed the entire physiologically relevant frequency spectrum of local field potentials and found a maximal peak of self-speech suppression at 100 Hz, as well as lower degrees of signal suppression in the theta range and in the overall power of the raw signal. The high degree of signal suppression at the 100 Hz band and adjacent frequencies provide evidence for a functional band of oscillatory activity. The suppressions observed in both the theta band and in the raw signal, which is dominated by power in low frequencies, are most likely due to the previously reported reduction in the N100 ERP component (Ford et al., 2001).

High gamma activity has been reported in a variety of functional modalities, including auditory- (Crone et al., 2001; Edwards et al., 2005; Trautner et al., 2006), motor- (Crone et al., 1998; Miller et al., 2007) and language- (Crone et al., 2001; Brown et al., 2008) related tasks. The γ_{High} response has been linked to neuronal firing rate and is believed to emerge from synchronous firing of neuronal populations (Mukamel et al., 2005; Liu and Newsome, 2006; Allen et al., 2007; Belitski et al., 2008; Ray et al., 2008; Cardin et al., 2009). The reduced γ_{High} responses we found in a substantial number of auditory electrodes during vocalization suggest a reduction in neuronal population activity in the underlying tissue. This attenuated response is consistent with previous reports of reduced auditory responses in non-invasive human studies as well as suppressed single unit responses (Ford et al., 2001; Eliades and Wang, 2003). Although all single unit studies report suppressed responses, there have been conflicting reports as to what proportion of the recorded neurons are suppressed during vocalization. Müller-Preuss et al. reported that over half the auditory neurons in the squirrel-monkey STG were suppressed, Eliades et al. found suppression in three quarters of marmoset-monkey STG neurons while Creutzfeldt et al. found only a minority of neurons suppressed in the human STG (Müller-Preuss and Ploog, 1981; Creutzfeldt et al., 1989; Eliades and Wang, 2003). Although the variability of these findings could be due to differences in species, our results suggest that it could be due to sampling of neurons from different regions of STG. The differential degree of suppression we observed in adjacent electrodes could be the direct result of averaging the activity of neuronal populations with different proportions of suppressed vs. non-suppressed neurons. Most of the auditory electrodes we recorded exhibited some degree of suppression, thus averaging these regions would result in an overall suppression as reported by non-invasive studies (Ford et al., 2001; Houde et al., 2002).

Auditory cortex suppression during vocalization has often been attributed to the influence of motor cortex. Eliades et al. have reported suppression of auditory neurons commencing as early as several hundred milliseconds prior to vocalization (Eliades and Wang, 2003). Nonetheless, there is no direct evidence in human or non-human primates linking this

suppression to motor cortex activity. We examined phase-locking and coherence measures between auditory electrodes and other regions of cortex including motor, pre-motor and frontal electrodes but observed no consistent coupling pattern. Similarly, Towle et al. were unable to find phase locking with motor regions (Towle et al., 2008). In both these studies the electrode coverage was limited to a 1-cm spacing over the lateral surface of the STG without direct recordings from primary auditory cortex. Although our results do not exclude motor cortex as the source of suppression they suggest a possible alternate model wherein the neuronal architecture of the auditory cortex itself supports suppression of self-generated speech through local cortico-cortical interactions.

Current theories support a forward model whereby corollary discharge signals from motor cortex, representing a prediction of impending acoustic input, modulate auditory cortex activity (Ford et al., 2001; Houde et al., 2002; Heinks-Maldonado et al., 2005). The theoretical framework for this model is based on work in the visual domain where an efferent copy of a motor command may be used to predict its sensory outcome (Sperry, 1950; Von Holst and Mittelstaedt, 1950). Evidence supporting this model in the auditory domain include two major findings: 1) The auditory cortex is mostly suppressed during vocalization; 2) Altering the expected auditory feedback abolishes auditory suppression (Curio et al., 2000; Ford et al., 2001; Houde et al., 2002; Eliades and Wang, 2003; Eliades and Wang, 2008). The source of the suppression remains unknown although it is widely assumed to originate in motor or pre-motor cortex. Our results provide evidence of differential levels of suppression as well as excited responses suggesting that auditory cortex isn't homogenously suppressed by a remote cortical source. An alternate possibility is that speech production shifts the auditory cortex to a different processing state (resulting from a global signal or a corollary discharge) where some sub-regions are suppressed, some excited and some remain unchanged. While it is possible that some sub-regions are directly attenuated by a remote source, the suppression and excitation might also be internally produced by the neuronal architecture of the auditory cortex.

Our current single-trial results provide evidence for stable responses within subjects across trials. Although different electrodes exhibit differences in the degree of suppression, these responses are remarkably consistent across trials. This suggests that every time we produce speech the auditory cortex responds with a specific pattern of suppressed and non-suppressed activity. This clarifies previous results – auditory cortex is not merely statistically suppressed on average but is functionally suppressed in a specific topographical pattern.

The human auditory cortex appears to have a specific topography of self-speech suppression that is stable across time, suggesting an intertwined mosaic of neuronal populations with suppressed and non-suppressed auditory responses. During vocalization the averaged activity of these populations exhibit a stable spatial pattern of varying suppression on the surface of the cortex, which is recorded as an averaged suppressed response from scalp electrodes. Our results complement both single-unit and non-invasive studies by offering an intermediate level of recording with a strong SNR, providing both temporal and spatial information. Furthermore, our data with higher density electrode recordings and other recent observations in humans (Chang et al., 2010; Flinker et al., 2010) provides evidence of independent auditory responses at 4-mm spacing. This observation suggests that the current typical ECoG sampling with 1-cm resolution is insufficient since it is averaging over smaller sub-regions of auditory cortex with potentially different response types.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Allen EA, Pasley BN, Duong T, Freeman RD. Transcranial magnetic stimulation elicits coupled neural and hemodynamic consequences. Science. 2007; 317:1918–1921. [PubMed: 17901333]

Bédard C, Kröger H, Destexhe A. Does the 1/f Frequency Scaling of Brain Signals Reflect Self-Organized Critical States? Phys Rev Lett. 2006; 97:4.

Belitski A, Gretton A, Magri C, Murayama Y, Montemurro MA, Logothetis N, Panzeri S. Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. Journal of Neuroscience. 2008; 28:5696–5709. [PubMed: 18509031]

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995; 57:289–300.

Brown EC, Rothermel R, Nishida M, Juhász C, Muzik O, Hoechstetter K, Sood S, Chugani HT, Asano E. In-Vivo Animation of Auditory-Language-Induced Gamma-Oscillations in Children with Intractable Focal Epilepsy. NeuroImage. 2008; 41:1120. [PubMed: 18455440]

Bruns A. Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? Journal of Neuroscience Methods. 2004; 137:321–332. [PubMed: 15262077]

Canolty R, Soltani M, Dalal SS, Edwards E, Dronkers NF, Nagarajan SS, Kirsch HE, Barbaro NM, Knight RT. Spatiotemporal dynamics of word processing in the human brain. Front Neurosci. 2007; 1:185–196. [PubMed: 18982128]

Cardin JA, Carlén M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai LH, Moore CI. Driving fast-spiking cells induces gamma rhythm and controls sensory responses. Nature. 2009; 459:663–667. [PubMed: 19396156]

Chang EF, Reiger JW, Johnson K, Berger MS, Barbaro NM, Knight RT. Categorical representation of phonemes in the human superior temporal gyrus. Nature Neuroscience. 2010 accepted.

Christoffels IK, Formisano E, Schiller NO. Neural correlates of verbal feedback processing: an fMRI study employing overt speech. Human Brain Mapping. 2007; 28:868–879. [PubMed: 17266104]

Creutzfeldt O, Ojemann, Lettich. Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. Exp Brain Res. 1989; 77:476–489. [PubMed: 2806442]

Crone NE, Miglioretti DL, Gordon B, Lesser RP. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain. 1998; 121(Pt 12):2301–2315. [PubMed: 9874481]

Crone NE, Hao L, Hart J, Boatman D, Lesser RP, Irizarry R, Gordon B. Electrocorticographic gamma activity during word production in spoken and sign language. Neurology. 2001; 57:2045–2053. [PubMed: 11739824]

Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. Human Brain Mapping. 2000; 9:183–191. [PubMed: 10770228]

Edwards E, Soltani M, Deouell LY, Berger MS, Knight RT. High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. Journal of Neurophysiology. 2005; 94:4269–4280. [PubMed: 16093343]

Eliades SJ, Wang X. Dynamics of Auditory-Vocal Interaction in Monkey Auditory Cortex. Cerebral Cortex. 2005; 15:1510–1523. [PubMed: 15689521]

Eliades SJ, Wang X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature. 2008; 453:1102–1106. [PubMed: 18454135]

Eliades SJ, Wang X. Sensory-Motor Interaction in the Primate Auditory Cortex During Self-Initiated Vocalizations. Journal of Neurophysiology. 2003; 89:2194. [PubMed: 12612021]

Flinker A, Chang EF, Barbaro NM, Berger MS, Knight RT. Sub- centimeter language organization in the human temporal lobe. Brain Lang. 2010 Advance online publication. doi: 10.1016/j.bandl.2010.09.009. PMID: 20961611.

Ford JM, Mathalon DH, Heinks T, Kalba S, Faustman WO, Roth WT. Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. The American journal of psychiatry. 2001; 158:2069–2071. [PubMed: 11729029]

Freeman W. Origin, structure, and role of background EEG activity. Part 1. Analytic amplitude. Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology. 2004; 115:2077–2088. [PubMed: 15294210]

Fu, MJ.; Daly, JJ.; Cavuşoğlu, MC. A detection scheme for frontalis and temporalis muscle EMG contamination of EEG data; Conf Proc IEEE Eng Med Biol Soc; 2006; p. 4514-4518.

Goncharova II, McFarland DJ, Vaughan TM, Wolpaw JR. EMG contamination of EEG: spectral and topographical characteristics. Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology. 2003; 114:1580–1593. [PubMed: 12948787]

Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM. Fine-tuning of auditory cortex during speech production. Psychophysiology. 2005; 42:180–190. [PubMed: 15787855]

Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. Journal of cognitive neuroscience. 2002; 14:1125–1138. [PubMed: 12495520]

Lane H, Tranel B. The Lombard Sign and the Role of Hearing in Speech. J Speech Hear Res. 1971; 14:677–709.

Levelt WJM. Monitoring and self-repair in speech. Cognition. 1983; 14:41–104. [PubMed: 6685011]

Liu J, Newsome WT. Local field potential in cortical area MT: stimulus tuning and behavioral correlations. Journal of Neuroscience. 2006; 26:7779–7790. [PubMed: 16870724]

Miller KJ, Leuthardt EC, Schalk G, Rao RP, Anderson NR, Moran DW, Miller JW, Ojemann JG. Spectral changes in cortical surface potentials during motor movement. Journal of Neuroscience. 2007; 27:2424–2432. [PubMed: 17329441]

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; 309:951–954. [PubMed: 16081741]

Müller-Preuss P, Ploog D. Inhibition of auditory cortical neurons during phonation. Brain Research. 1981; 215:61–76. [PubMed: 7260601]

Numminen J, Salmelin R, Hari R. Subject's own speech reduces reactivity of the human auditory cortex. Neuroscience Letters. 1999; 265:119–122. [PubMed: 10327183]

Nunez, PL. Electric Fields of the Brain: The Neurophysics of EEG. 2nd Edition. Oxford University Press; New York: 2005. SR

Pritchard W. The Brain in Fractal Time: 1/F-Like Power Spectrum Scaling of the Human Electroencephalogram. Int J Neurosci. 1992; 66:119–129. [PubMed: 1304564]

Ray S, Crone NE, Niebur E, Franaszczuk PJ, Hsiao SS. Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography. Journal of Neuroscience. 2008; 28:11526–11536. [PubMed: 18987189]

Sohal VS, Zhang F, Yizhar O, Deisseroth K. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. Nature. 2009; 459:698. [PubMed: 19396159]

Sperry R. Neural basis of the spontaneous optokinetic response produced by visual inversion. Journal of comparative and physiological psychology. 1950; 43:482–489. [PubMed: 14794830]

Towle VL, Yoon H, Castelle M, Edgar JC, Biassou NM, Frim DM, Spire J, Kohrman MH. ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain. 2008; 131:2013. [PubMed: 18669510]

Trautner P, Rosburg T, Dietl T, Fell J, Korzyukov OA, Kurthen M, Schaller C, Elger CE, Boutros NN. Sensory gating of auditory evoked and induced gamma band activity in intracranial recordings. NeuroImage. 2006; 32:790–798. [PubMed: 16809054]

Von Holst E, Mittelstaedt M. Das Reafferenzprinzip. Naturwissenschaften. 1950; 37:464–476.

Whitham EM, Lewis T, Pope KJ, Fitzgibbon SP, Clark CR, Loveless S, DeLosAngeles D, Wallace AK, Broberg M, Willoughby JO. Thinking activates EMG in scalp electrical recordings. Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology. 2008; 119:1166–1175. [PubMed: 18329954]

Wise RJS, Greene J, Buchel C, Scott SK. Brain regions involved in articulation. Lancet. 1999:1057–1061. [PubMed: 10199354]

Yates A. Delayed auditory feedback. Psychol Bull. 1963; 60:213–232. [PubMed: 14002534]

Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY. Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron. 2008; 58:429–441. [PubMed: 18466752]
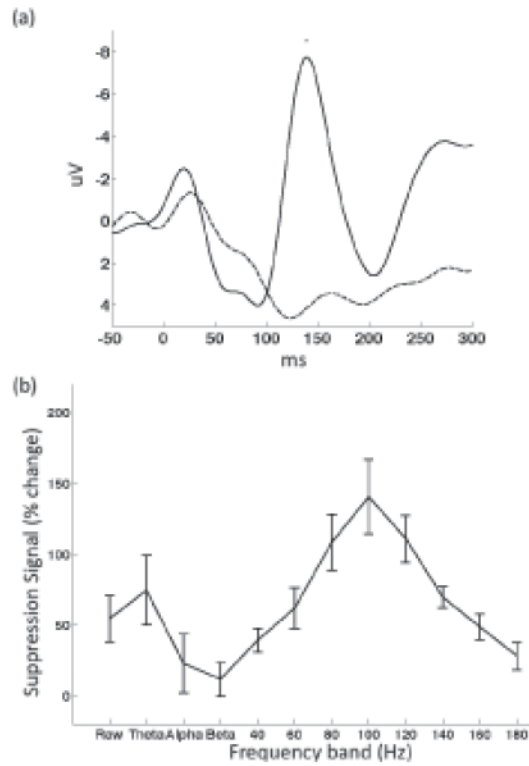
Figure 1.
(a) Auditory event related potentials locked to hearing (solid) and speaking (dashed) vowels.
(b) Difference signal between hearing and speaking compared to baseline across the different frequency bands. Error bars represent the SEM across subjects.
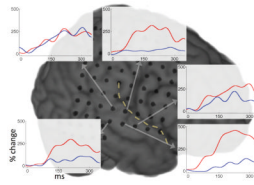
Figure 2.
Averaged γ_{High} power traces locked to hearing (red) and speaking (blue) vowels in five different electrodes across the STG of subject S3.
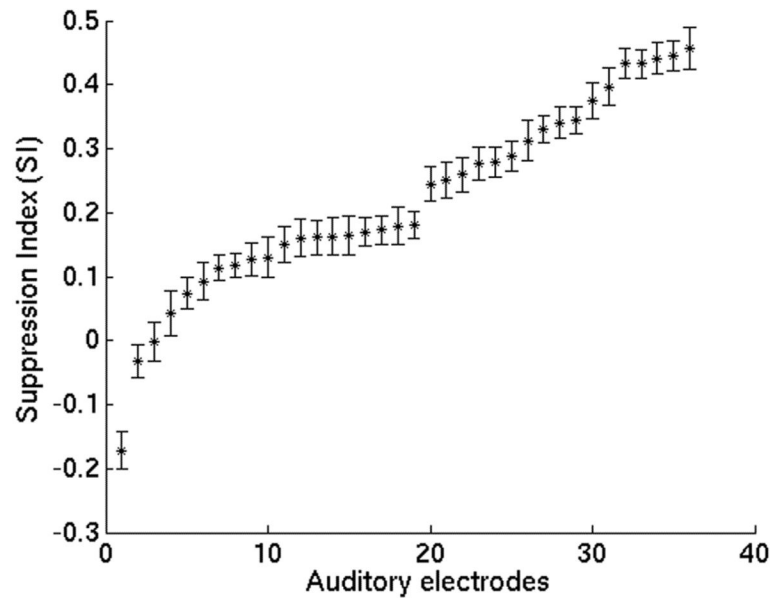
Figure 3.
Electrode Suppression Indices (SI) for all 36 γ_{High} auditory electrodes sampled from all subjects. SI = $(P_{hear} - P_{speak})/(P_{hear} + P_{speak})$, where P denotes event-related γ_{High} power.
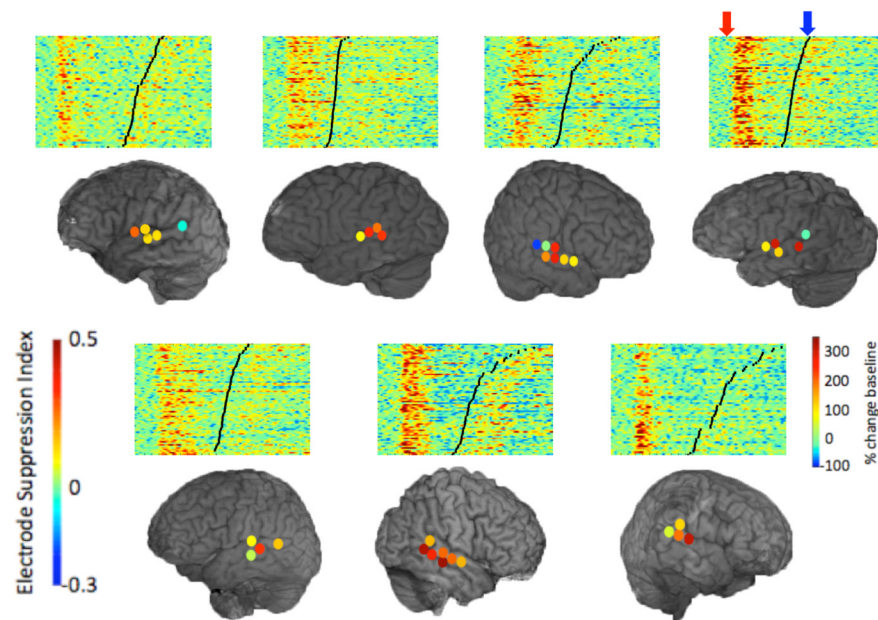
Figure 4.
Regional suppression topography of all the γ_{High} auditory electrodes in each subject. Colored dots represent suppression indices in each electrode. Above each subject are vertically stacked single trial γ_{High} traces shown for a representative electrode from each subject. Single trial traces are locked to hearing stimuli (red arrow) and black lines denote speech onset (blue arrow).
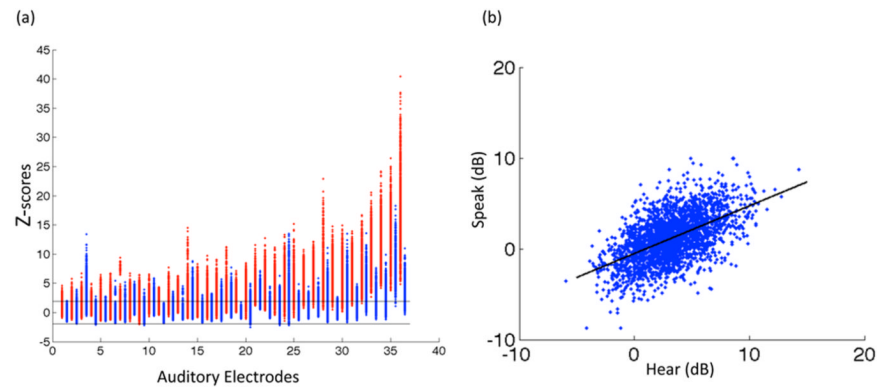
Figure 5.
(a) Bootstrapped hearing (red) and speaking (blue) events for all γ_{High} auditory electrodes compared with a baseline distribution (p<0.05 significance levels marked by black horizontal lines). Electrodes are sorted according to the mean of hearing events. (b) Single trial γ_{High} power values for hearing versus speaking across all electrodes.
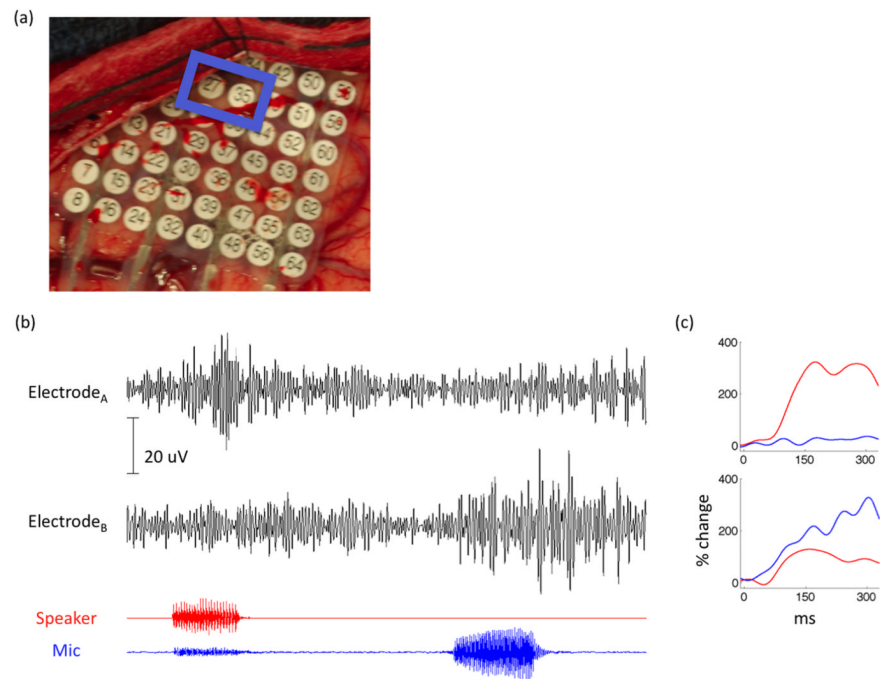
Figure 6.
(a) Intraoperative image of a high-density 4 mm electrode grid. (b) High frequency oscillatory responses are shown for two adjacent electrodes during a 2 second epoch (raw traces filtered at 70-150 Hz). Electrode A responds selectively to external stimuli (speaker plotted in red) while Electrode B responds predominantly to self-produced speech (microphone plotted in blue). (c) Averaged event related power changes locked to hearing (red) and speaking (blue) stimuli in the two adjacent electrodes.