Title
The electrophysiology of language perception and production

Author
Flinker, Adeen

Publication Date
2012

Peer reviewed|Thesis/dissertation

The Electrophysiology of Language Perception and Production

By
Adeen Max Flinker

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Neuroscience
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Robert T. Knight, Chair
Professor Frederic E. Theunissen
Professor Shaowen Bao
Professor George Lakoff

Spring 2012

Abstract

The electrophysiology of language perception and production

by

Adeen Max Flinker

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Robert T. Knight, Chair

For over a century, an abundance of research has tried to elucidate the neurobiological basis of language processing in the human cortex. Neuroimaging and lesion studies have provided great insight into what functions different brain structures subserve. While these techniques provide a high spatial resolution they are limited in the temporal domain. Conversely, contributions from non-invasive electrophysiology provided a high temporal resolution with a limited ability to localize cortical sources. The combined spatial and temporal dynamics of cortical processing during language perception and production remains largely unknown. This dissertation addresses this issue by employing unique neuronal population recordings from neurosurgical patients performing linguistic tasks. The studies described here elucidate the timing, magnitude and spatial extent of cortical processing during perception and production of language. The results provide evidence on the level of single-trial that: 1) A rich network of independent and spatially distinct functional sub-regions of cortex subserve perception and production of language. 2) Neighboring sub-regions 4 mm apart can exhibit inverse functional specific responses to linguistic stimuli and self produced speech. 3) Broca's area is not involved in the actual act of articulation but rather in speech preparation and interfacing perception and production. Taken together, these results defy century old dogmas and suggest that language is supported by a complex network of independent sub-regions, with Broca's area acting as a mediator between perception and production rather than as the seat of articulation.

Table of Contents

This dissertation is dedicated to my grandmother, Klara Haya Weisman,
who knew much suffering but spread only love

Acknowledgments

This dissertation represents one of the most exciting periods of my life thus far. I could not have finished the work without the support, consult and love of my family and friends. I would like to thank my father, mother and brother without whom life would not be the same loving place and my family and friends in Israel who supported my continued life abroad. During my time as a graduate student I experienced the joy of meeting the love of my life, Carmela, who stood by me every step of the way – toda ahuvaty. I would like to thank all my friends I met during my wondrous time in Berkeley who were with me during work and play. I would also like to thank my advisor, Prof. Robert T. Knight, who offered his guidance and support as well as infinite academic resources and freedom to accomplish what would have been impossible without his help.

Last but not least, this work could not have been possible without the consent and cooperation of the patients undergoing neurosurgical treatment.

The studies in this dissertation were supported in part by an NIH training fellowship F31NS065656

Chapter 1

Introduction

The classical model of language is mainly based on clinical research of brain-injured aphasic patients in the 19th century. Pioneers such as Paul Broca and Carl Wernicke described patients with severe expressive deficits (non-fluent and agrammatic speech output) in the face of relatively good comprehension and patients with profound receptive deficits (auditory language comprehension) in the face of fluent speech output. These cognitive impairments were correlated with brain lesions described post mortem. Broca's aphasia (expressive) was correlated with an anterior brain region called Broca's area (BA 44/45) and Wernicke's aphasia (receptive) was correlated with a posterior brain region called Wernicke's area (BA 22). The model, revived in the 20th century by Norman Geschwind, posited a temporal frontal network with a temporal region (Wernicke's area) responsible for comprehension connected via the arcuate fasciculus to a frontal region (Broca's area) responsible for articulation (Broca, 1861; Wernicke, 1874; Geschwind, 1970).

Over the past century lesion studies have shown that this model is insufficient in explaining the wide array of aphasic symptoms and is anatomically underspecified. Patients with both anterior and posterior aphasias were found to make errors in receptive phoneme identification and discrimination tasks (Blumstein et al., 1977a; Blumstein et al., 1977b). Moreover, lesions involved in different aphasias typically cover more extensive temporal or frontal regions and are not limited to the classical language areas, complicating structure-function correlations (Damasio, 1992; Dronkers et al., 2004). A striking limitation of the classical model is evident in lesion studies showing that cortical damage limited to Broca's area does not cause a Broca's aphasia but rather results in a transient, rapidly improving mutism (Mohr et al., 1978), and that articulation deficits do not necessarily involve damage to Broca's area (Dronkers, 1996).

The advent of modern neuroimaging techniques propelled an explosion of research aimed at elucidating language processing in normal subjects. In accordance with the classical theory of language, early PET studies confirmed the role of the superior temporal gyrus in receptive language function (Peterson et al., 1989; Zatorre et al., 1996; Demonet et al., 1992; Price et al., 1996). Recent evidence suggests a more complex network of cortical structures supporting language (Hickok and Poeppel, 2007; Scott and Johnsrude, 2003). Furthermore, neuroimaging studies have shown consistent Broca's area activation in receptive tasks that do not require overt articulation (Zatorre et al., 1992; Price et al., 1996; Binder et al., 1997). These neuroimaging studies typically report activation sites spanning several centimeters of cortex and represent an average across several subjects. Conversely, intraoperative language mapping using electrical cortical stimulation (ESM) report a high degree of inter-subject variability in the location of cortical language sites (Ojemann et al., 1989; Sanai et al. 2008). This inter-subject variability suggests that activation maps currently drawn from neuroimaging data are potentially obscuring a finer grain cortical organization of language.

While the aforementioned techniques possess a high spatial resolution they lack the millisecond temporal resolution necessary to track language processing which unfolds in the sub-second time domain. Noninvasive electrophysiology in humans provides excellent temporal resolution and has been available for over 80 years (Berger, 1929). A clear electrical waveform signature evoked by an event or stimulus requires averaging of multiple events and is typically averaged across subjects. One of the earliest evoked potentials identified was a positive-negative electrical wave elicited by sound stimuli (Davis, 1939). The first negative peak of the wave is called the N1 or N100 and peaks at a latency of about 100 ms. The N1 component marks one of the first steps in cortical auditory processing being generated just after the middle latency auditory response at 20-50 ms (Näätänen et al., 1987). Later processing stages specific to language processing have been identified, indexing syntactic errors (LAN, ELAN, P600: Neville et al., 1991; Osterhout and Holcomb, 1992; Friederici, 1993; Hahne and Friederici, 1999; Kaan et al., 2000) as well as semantic mismatch (N400: Kutas and Hillyard, 1980; Kutas and Federmeier, 2000). Nevertheless, non-invasive electrophysiology is limited in localization of cortical sources. As a consequence, the timing of activity across cortex during language processing has been mostly pieced together from neuroimaging and non-invasive electrophysiological studies (Friederici, 2002; Indefery and Levelt, 2004).

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 and Srinivasan, 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, electrocorticographic (ECoG) recordings largely circumvent these issues providing high a signal to noise ratio (SNR), physical sampling from well-defined cortical sources as well as access to richer spectral content. A recently described high gamma band (γ_{High}: 70-150 Hz) provides a reliable index of cortical activation and increased spike activity that is not readily seen in scalp EEG (Crone et al., 1998; Crone et al., 2001a; Crone et al., 2001b; Canolty et al., 2006; Allen et al., 2007; Sohal et al., 2009). The outstanding signal quality afforded by these recordings is ideal for investigating the spatio-temporal dynamics of cortical language processing.

The perception and production of language is innately human. Nevertheless, several non-human species possess the ability to perceive and produce vocalizations providing an opportunity to investigate mechanisms of perception and production. Evidence from non-human primates supports a hierarchal organization of functionally distinct subdivisions of auditory cortex subserving auditory perception (Hackett et al., 1998; Rauschecker, 1998). While these pathways have served as the basis for inference to the human system, many discrepancies still remain regarding the functional organization of human speech perception (Scott et al., 2000; Wise et al., 2001; Hickok and Poeppel, 2007). Similarly, there have been discrepancies in reports between single-unit studies and non-invasive human electrophysiology regarding how the auditory system processes self-produced vocalizations (Müller-Preuss and Ploog, 1981; Ford, 2001; Eliades and Wang, 2008). The aim of this dissertation is to elucidate the spatiotemporal dynamics of

phoneme and word processing during perception and production of speech. The studies described here employ rare neurosurgical recordings in order to bridge results from the animal and human cortical systems and elucidate the spatiotemporal dynamics of language processing in the human cortex.

Chapter 2

Temporal lobe activity during perception and production

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.

2.1 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, 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., 2001b; 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.

2.2 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 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.

	Age	Sex	Craniotomy	Pathology
S1	18	Male	Left hemisphere	Tuberous sclerosis
S3	12	Male	Right hemisphere	Right posterior frontal encephalomalacia
S4	18	Male	Left hemisphere	Mesial temporal sclerosis
S5	34	Male	Left hemisphere	Cortical heterotopia
S7	51	Female	Right hemisphere	Right parietal stroke

Table 1
Details of subjects with specific histological pathologies

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_{High} 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 -> 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 -> -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_{High} 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 -> 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->-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 -> 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).

2.3 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 (red trace) and this response is severely reduced for the speaking ERP (blue 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, $p<0.05$).

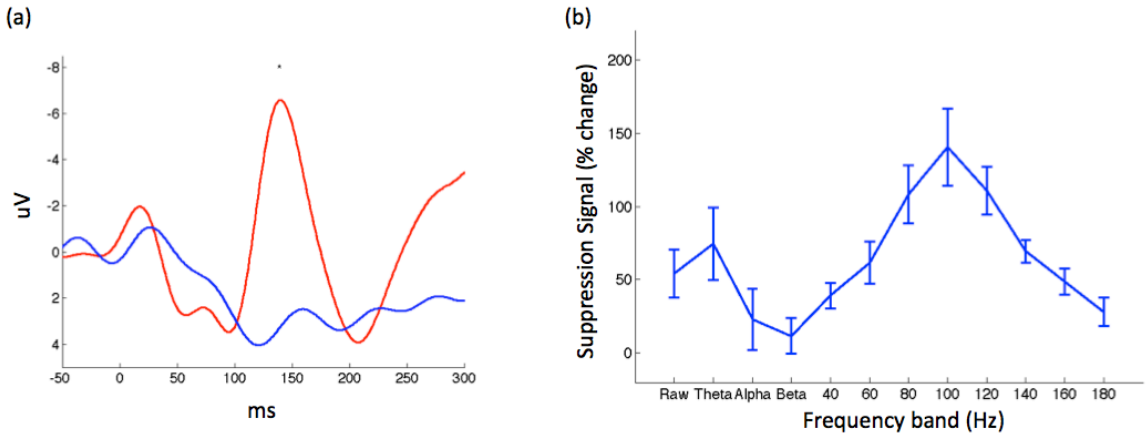


Figure 1

(a) Auditory event related potentials locked to hearing (red) and speaking (blue) vowels.
(b) Difference signal between hearing and speaking compared to baseline across the different frequency bands. Error bars represent the SEM across subjects.

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).

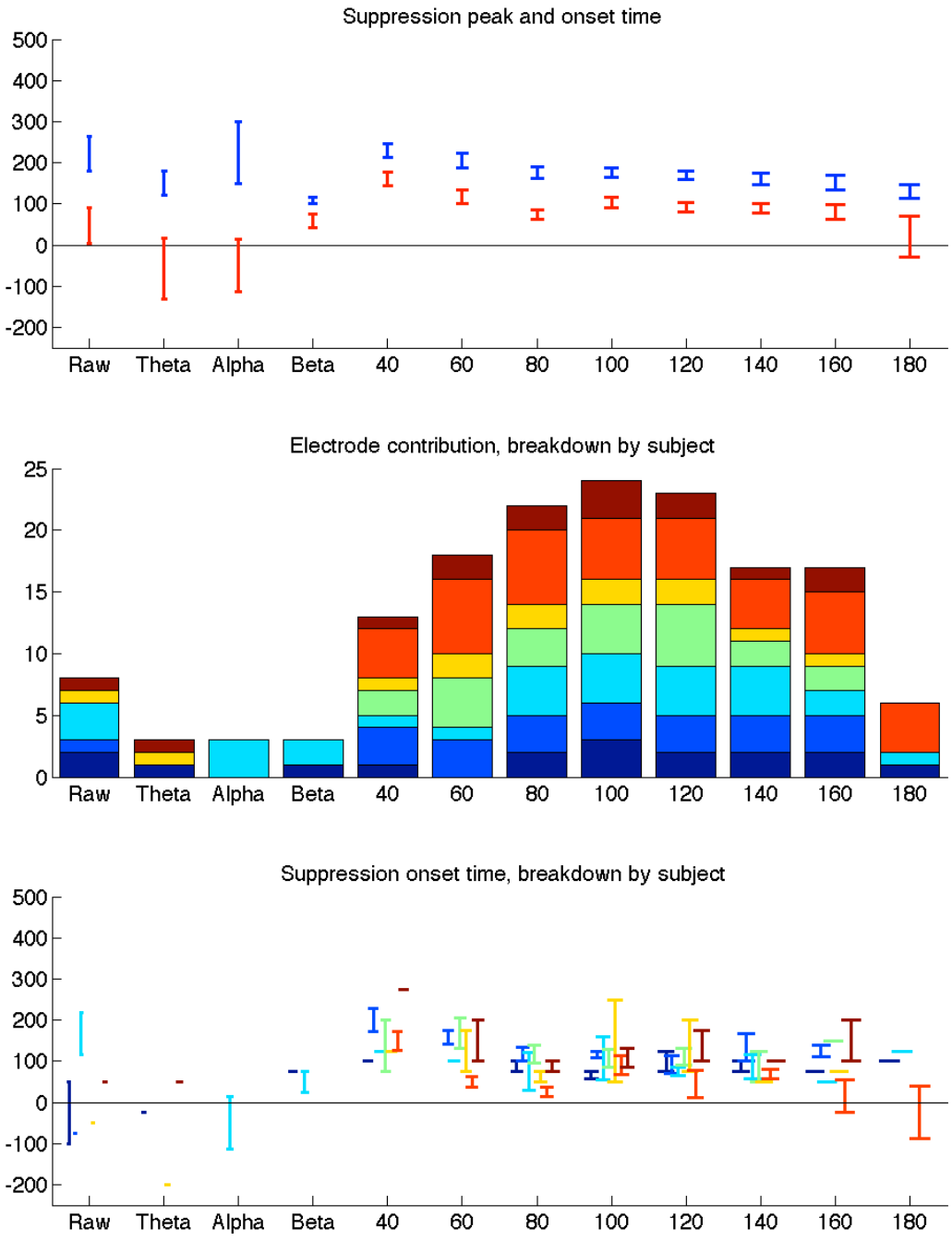


Figure S1

Suppression peak and onset time data. (a) Mean suppression onset time (red) and peak (blue) for all electrodes across all subjects exhibiting suppression in a specific frequency band. (b) Histogram of the number of electrodes with significant suppression in each frequency band, broken down by subject (colormap represents each individual subject: dark blue, blue, light blue, green, yellow, red, dark red represent subjects 1 through 7 respectively). (c) Mean suppression onset time data shown for each subject individually (colormap represents each individual subject: dark blue, blue, light blue, green, yellow, red, dark red represent subjects 1 through 7 respectively).

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).

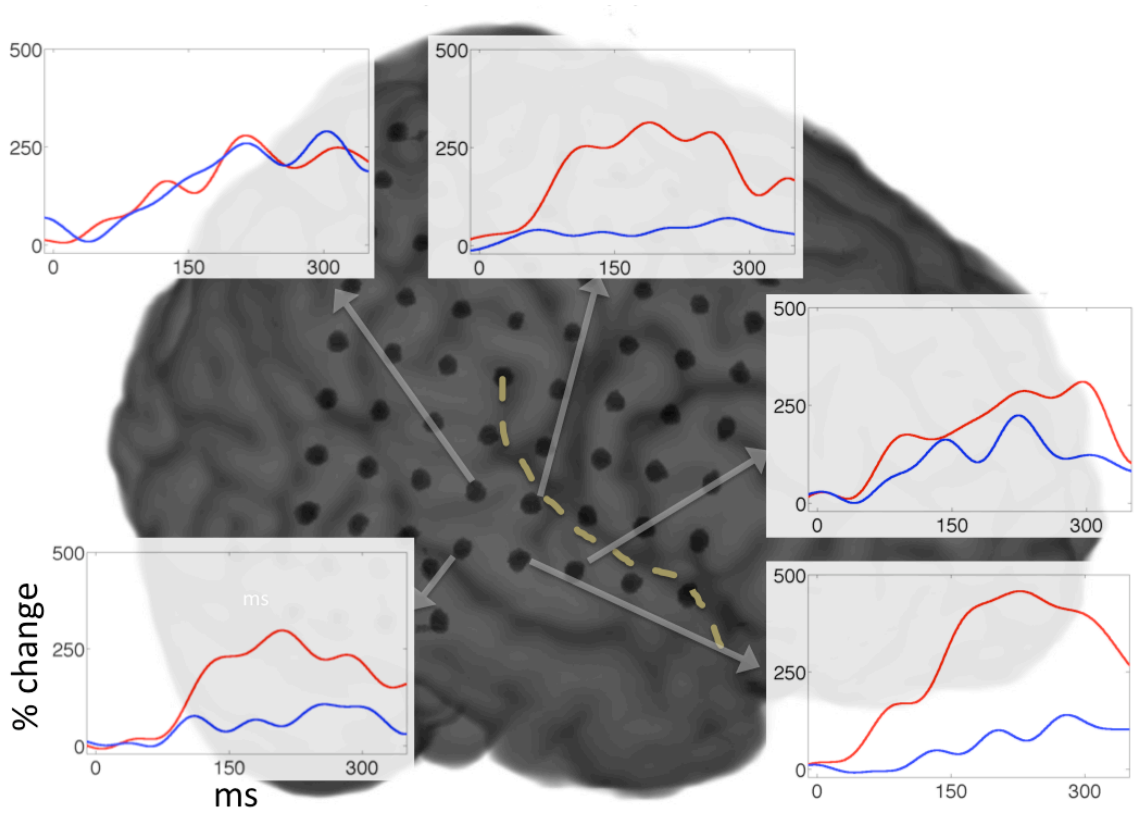


Figure 2

Averaged γ_{High} power traces locked to hearing (red) and speaking (blue) vowels in five different electrodes across the STG of subject S3.

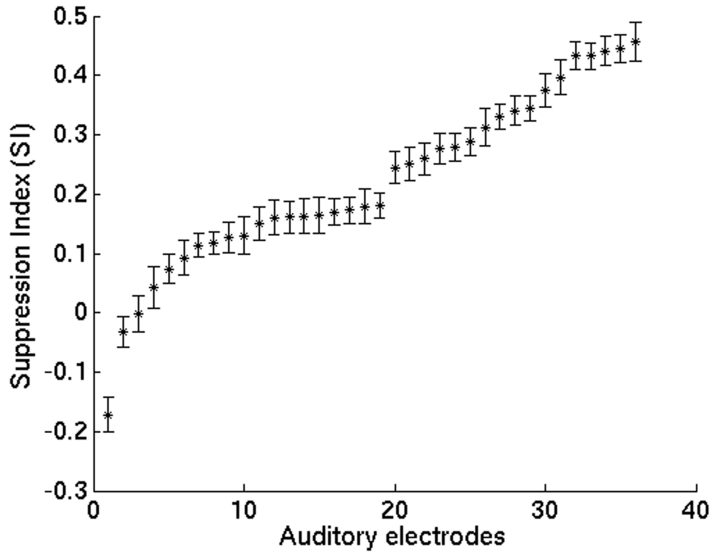


Figure 3

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.

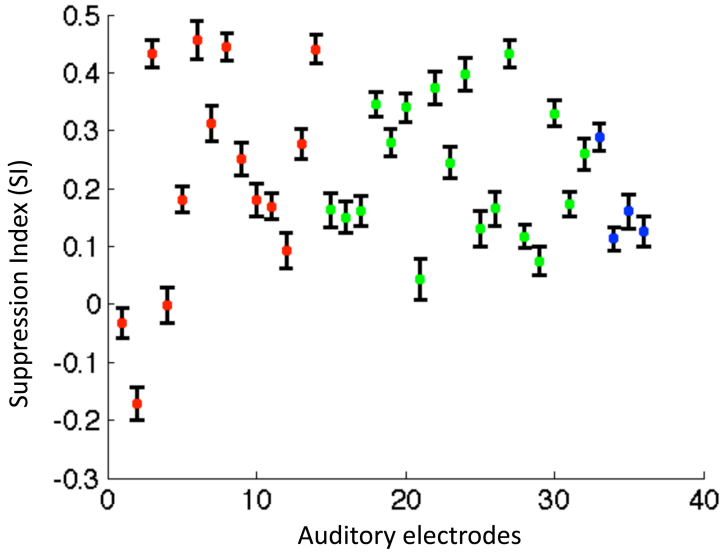


Figure S2

Electrode Suppression Indices (SI) for all 36 γ_{High} auditory electrodes across all the subjects. Electrodes are grouped according to anatomical location relative to the lateral surface of Heschl Gyrus: posterior (red), central (green) and anterior (blue).

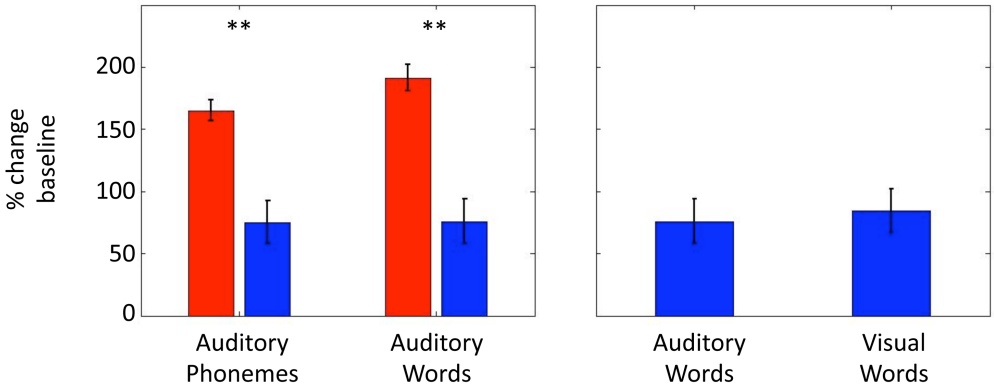


Figure S3

γ_{High} spectral responses averaged across three subjects who performed a phoneme repetition task, a word repetition task and a word reading task. On the left γ_{High} power responses are shown for hearing (red) and speaking (blue) auditory stimuli. γ_{High} responses locked to hearing and speaking auditory stimuli were significantly different in both words and phonemes (t-test, $p<0.01$). γ_{High} responses to hearing phonemes did not differ significantly from hearing words (Wilcoxon rank-sum, $p>0.05$). On the right γ_{High} power responses are shown locked to production of words during an auditory repetition task and a visual reading task. γ_{High} responses locked to production of auditory and visual stimuli did not differ (Wilcoxon rank-sum, $p>0.05$).

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 S4).

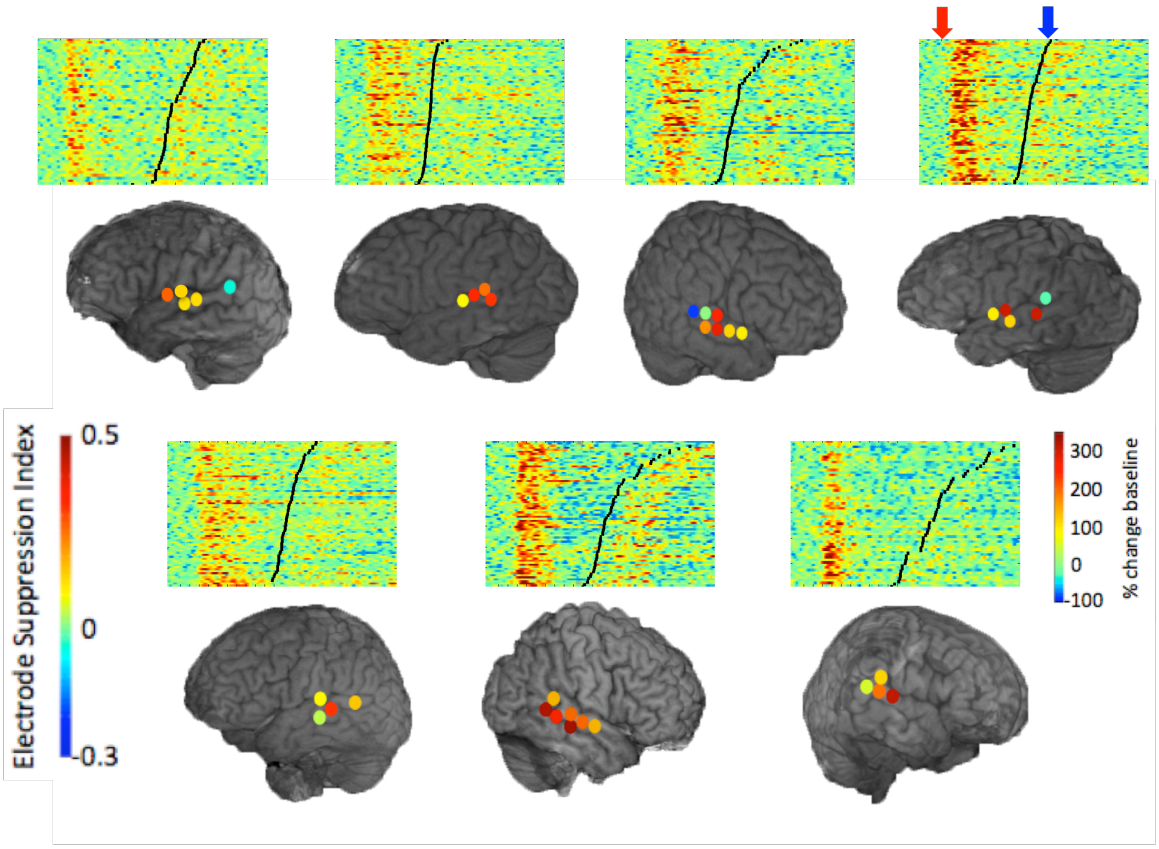


Figure 4

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).

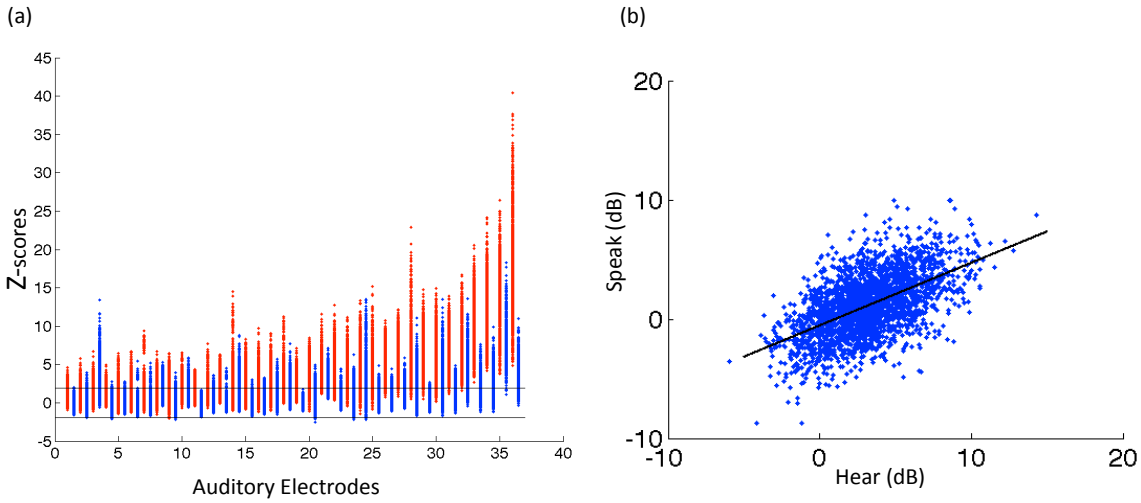


Figure 5

(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.

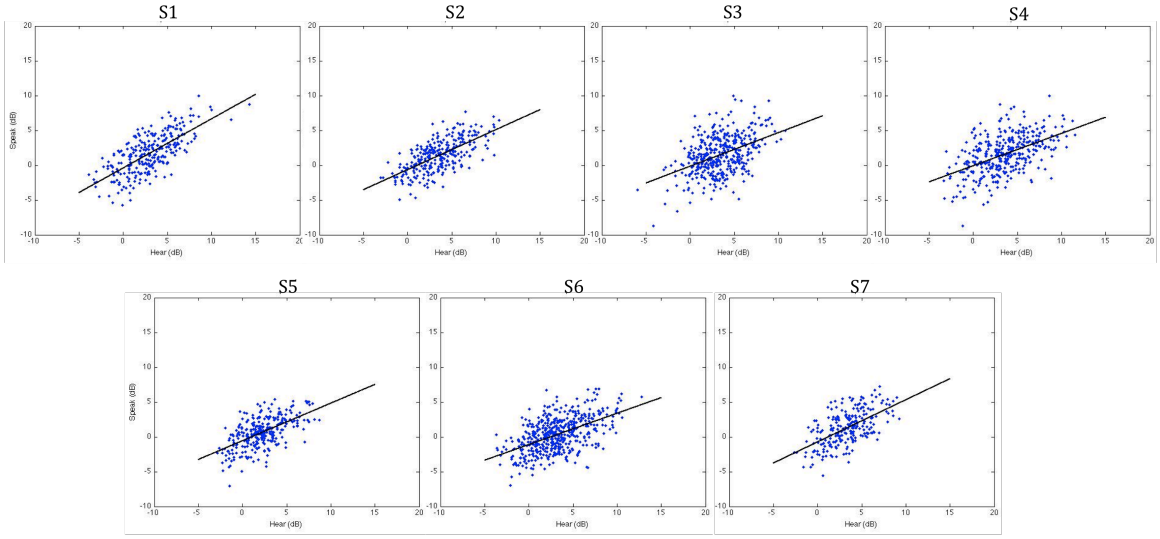


Figure S4

Single trial γ_{High} power values for hearing versus speaking across electrodes in each individual subject. All linear relationships were significant ($p<0.001$; R = 0.70, 0.66, 0.44, 0.50, 0.59, 0.53, 0.59 for subjects S1-S7 respectively).

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)).

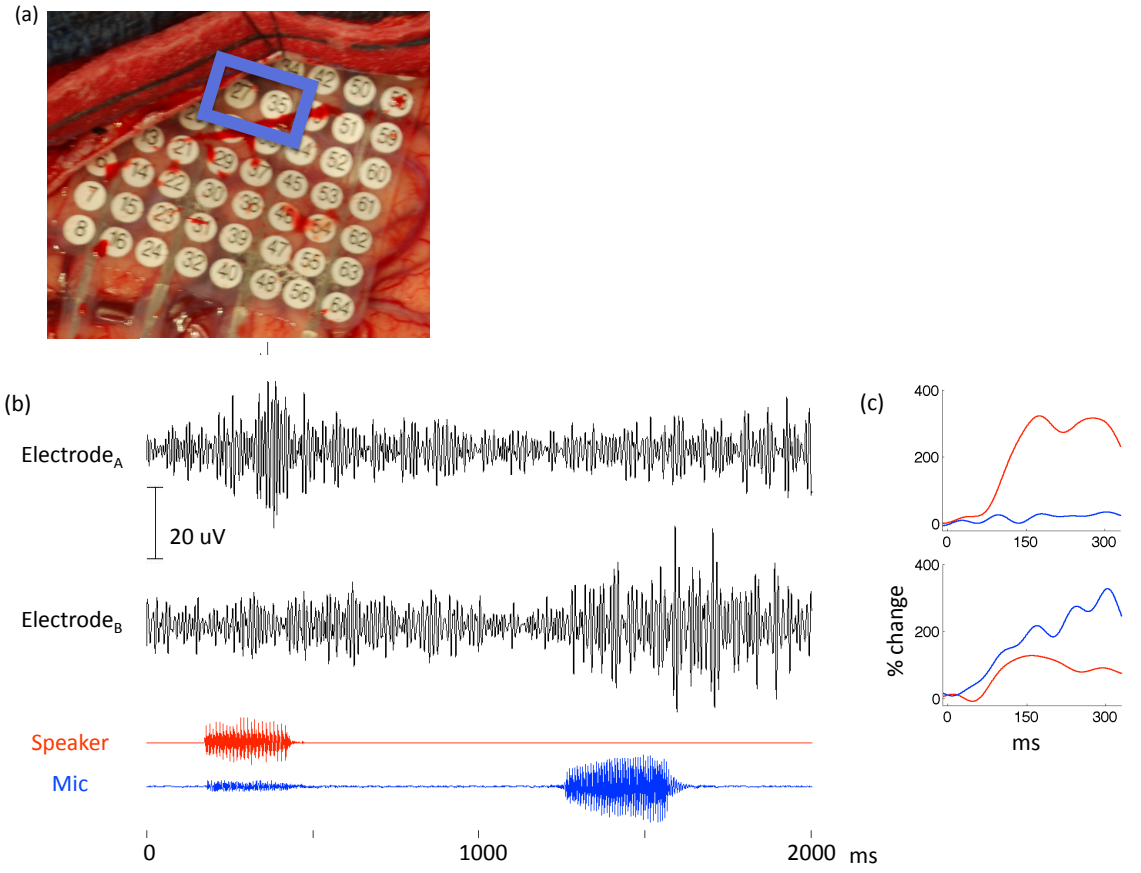


Figure 6

(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.

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., 2001b; 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., 2001b; Edwards et al., 2005; Trautner et al., 2006), motor- (Crone et al., 1998; Miller et al., 2007) and language- (Crone et al., 2001b; 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., 2011) 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.

Chapter 3

Sub-centimeter functional responses in the temporal lobe

Abstract

The human temporal lobe is well known to be critical for language comprehension. Previous physiological research has focused mainly on noninvasive neuroimaging and electrophysiological techniques with each approach requiring averaging across many trials and subjects. The results of these studies have implicated extended anatomical regions in peri-sylvian cortex in speech perception. These non-invasive studies typically report a spatially homogenous functional pattern of activity across several centimeters of cortex. We examined the spatiotemporal dynamics of word processing using electrophysiological signals acquired from high-density electrode arrays (4 mm spacing) placed directly on the human temporal lobe. Electrocorticographic (ECoG) activity revealed a rich mosaic of language activity, which was functionally distinct at four mm separation. Cortical sites responding specifically to word and not phoneme stimuli were surrounded by sites that responded to both words and phonemes. Other sub-regions of the temporal lobe responded robustly to self-produced speech and minimally to external stimuli while surrounding sites at 4 mm distance exhibited an inverse pattern of activation. These data provide evidence for temporal lobe specificity to words as well as self produced speech. Furthermore, the results provide evidence that cortical processing in the temporal lobe is not spatially homogenous over centimeters of cortex. Rather, language processing is supported by independent and spatially distinct functional sub-regions of cortex at a resolution of at least 4 mm.

Introduction

The role of the human temporal lobe has been vigorously studied since Carl Wernicke's first account of language comprehension deficits in stroke patients (Wernicke, 1874). Numerous neuroimaging studies have employed various paradigms in an attempt to elucidate the neuroanatomical pathways of language perception. Across these studies the superior temporal gyrus and superior temporal sulcus have been consistently implicated in the perception of speech (Binder et al., 1997; Démonet et al., 1992; Price et al., 1996; Wise et al., 1991; Zatorre et al., 1992).

Evidence from non-human primates supports a hierarchal organization of functionally distinct subdivisions of auditory cortex (Hackett et al., 1998; Rauschecker, 1998). Studies in humans support a similar sub-division of auditory cortex supporting distinct anterior and posterior streams of information processing (Scott et al., 2000; Wise et al., 2001; Hickok and Poeppel, 2007). Furthermore, Binder et al. identified different sub-regions of the superior temporal lobe activated in processing speech and non-speech stimuli (Binder et al., 2000). Nonetheless there remain discrepancies between the studies regarding the exact anatomical pathways as well as the functional significance of the different sub-divisions of auditory cortex. Scott et al. reported a posterior and anterior subdivision processing unintelligible and intelligible speech respectively (Scott et al., 2000). Wise et al. focused on two posterior subdivisions: the posterior STS processing perception and retrieval of words, and the medial temporoparietal junction processing speech production (Wise et al., 2001). Lastly, Hickok & Poeppel proposed a dual-stream model of speech processing including a ventral stream (superior middle temporal lobe) processing speech comprehension and a dorsal stream (posterior dorsal temporal lobe, parietal operculum and posterior frontal lobe) processing auditory-motor integration (Hickok and Poeppel, 2007).

Neuroimaging studies investigating the neuroanatomical functional organization of the superior temporal gyrus typically report activation sites spanning several centimeters of cortex and represent an average across several subjects (Wise et al., 1991; Démonet et al., 1992; Zatorre et al., 1992; Price et al., 1996; Binder et al., 1997). Conversely, intraoperative language mapping using electrical cortical stimulation (ESM) report a high degree of inter-subject variability in the location of cortical language sites using 1 cm resolution electrodes (Ojemann et al., 1989; Sanai et al., 2008). This inter-subject variability suggests that activation maps currently drawn from neuroimaging data are averaging over a distribution of cortical sub-regions involved in speech processing potentially obscuring a finer grain cortical organization of language. While the functional neuroanatomy of language is likely common across subjects, some cortical sites could have a spatially dense organization, which varies across subjects. That is, use of a common coordinate frame across subjects could blur differences in regional cytoarchitecture across individual subjects. In order to address the spatial distribution of cortical activity during word and phoneme processing we recorded intraoperative electrocorticographic (ECoG) activity directly from the surface of the human temporal lobe using high-density (4 mm spacing) electrode arrays. ECoG recordings acquired directly from cortex provide a rich electrophysiological signal with spectral content not

readily seen in conventional EEG. Spectral High Gamma band activity (γ_{High} >70 Hz) is an ideal index for cortical activity and has been reported to reliably track neuronal activity in various functional relevant modalities including, auditory- (Crone et al., 2001a; Edwards et al., 2005; Trautner et al., 2006), motor- (Crone et al., 1998; Miller et al., 2007) and language- (Crone et al., 2001b; Tanji et al., 2005; Canolty et al., 2007; Brown et al., 2008; Towle et al., 2008) related tasks. Furthermore, 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).

Methods

Subjects

Four subjects participated in the study at University of California San Francisco (UCSF) while undergoing intraoperative neurosurgical treatment for refractory epilepsy (Subjects S1 and S3) or tumor resection (Subjects S2 and S4). Treatment involved a surgical procedure including intraoperative awake language and motor mapping followed by tailored resection of the damaged tissue under ECoG guidance. After all clinical mapping was performed, the surgeon (EC) placed a high density electrode array with inter-electrode spacing of 4 mm. The subjects performed either passive listening or repetition tasks (Supplemental Table 1) for several minutes after which time the grid was removed and the surgeon continued clinical treatment. The subjects provided written and oral consent prior to the surgery and were informed that the task was for research purposes. All subjects were fluent in English and had no language deficits. During surgery, the subjects were informed by the surgeon when the clinical mapping was over and the research task was completed under the discretion of the surgeon. All medical treatment including the size and location of the craniotomy site were dictated solely by the clinical needs of the patient under the discretion of the surgeon. The study protocol was approved by the UC San Francisco and UC Berkeley Committees on Human Research.

Tasks and Stimuli

Four subjects (S1-4) participated in short five-minute tasks consisting of either passive listening (S1-2) or active repetition (S2-4) of phonemes or words (see Supplemental Table 1 for breakdown per subject). The phoneme passive listening task consisted of the synthesized syllables /ba/, /da/ and /ga/ digitized with a sampling rate of 20 KHz and 16 bit precision. Phoneme stimuli were presented with a jittered inter stimulus interval (ISI) of 1.5 seconds ±100 ms (random jitter), each stimulus lasted approximately 150 ms. Subjects were instructed that they were going to hear several speech syllables and they were asked to concentrate on whether they heard /ba/, /da/ or /ga/. The phoneme repetition task consisted of the synthesized syllables /ba/ and /pa/ digitized with a sampling rate of 20 KHz and 16 bit precision. Phoneme stimuli were presented with a jittered inter stimulus interval (ISI) of 3 seconds ±100 ms (random jitter), each stimulus lasted approximately 150 ms. Subjects were instructed to listen and then repeat aloud each speech syllable as best they could. The word passive listening task consisted of word stimuli, digitally prerecorded from a female native speaker of English,

acquired at a sampling rate of 44 KHz and 16 bit precision. Word stimuli consisted of 23 pseudo-words (3 phonemes in length), 23 real words (3 phonemes in length) and 4 proper names (5 phonemes in length). Real words were all high frequency (Kucera Francis log scale 2-2.4), both real and pseudo-words were controlled for phonotactic probabilities using the Irvine Phonotactic Online Dictionary (Vaden et al., 2009).
Word stimuli were presented with a jittered inter stimulus interval (ISI) of 2 seconds ±250 ms (random jitter), stimuli varied in length (400-700 ms) with a mean of 525 ms and standard deviation of 100 ms. Subjects were instructed that they were going to hear several words, some with meaning and some without, and they were asked to listen to what they heard. The word repetition task consisted of the same word stimuli described above. Word stimuli were presented with a jittered inter stimulus interval (ISI) of 4 seconds ±250 ms (random jitter). Subjects were instructed that they were going to hear several words, some with meaning and some without, and they were asked to repeat aloud each word they heard as best they could. All subjects were presented with the digital audio stimuli vowels via two speakers placed in front of them. During the repetition tasks vocalization responses were recorded by using a microphone placed in front of the patient. All peripheral equipment was placed in the same locations traditionally used by the neurology team who administer language mapping tasks. Both the speaker and microphone channels were fed directly to the recording system in order to record stimulus presentation and responses simultaneously with the electrophysiological signals.

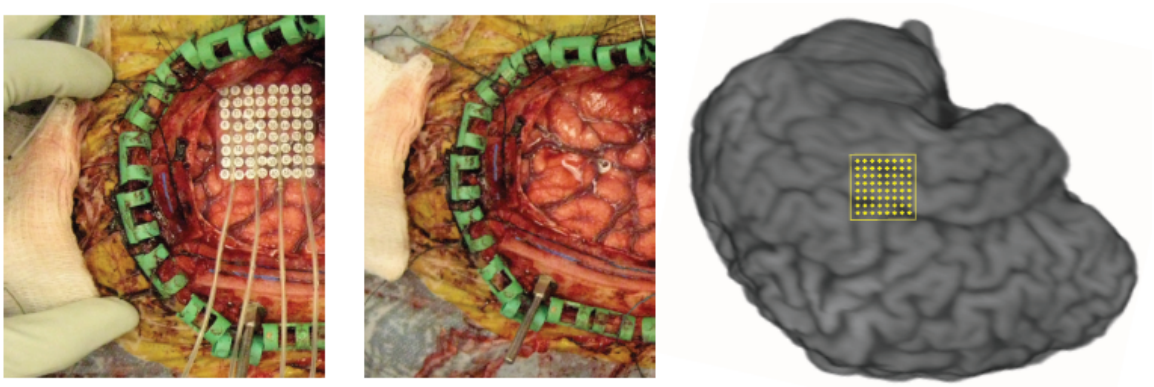
Subject	Passive Listening	Active Repetition	Age	Sex	Procedure
S1	Phoneme[1] Word		46	M	Temporal lobectomy
S2	Phoneme[1]	Phoneme[2] Word	53	M	Parietal lesionectomy
S3		Phoneme[2] Word	33	F	Temporal lobectomy
S4		Phoneme[2]	51	M	Frontal lesionectomy

Supplemental Table 1
Breakdown of stimuli and tasks preformed by each subject including Age, Sex and intraoperative procedure performed. Superscripts refer to the two types of phoneme stimuli described in the Methods section.

Electrode Placement and Selection

High density grids consisted of electrode arrays with 4 mm inter-electrode spacing manufactured by Ad-Tech Medical Instrument Corporation (http://www.adtechmedical.com/) assuring hospital-grade, FDA-approved standards. All electrode arrays were placed by the surgeon and covered the posterior lateral surface of the superior temporal gyrus. Supplemental Figure 4 shows an intraoperative image with and without the electrode grid alongside a reconstruction of the MRI. Grid locations were ascertained by the performing neurosurgeon using the intraoperative images as well as stereotactic coordinates. Word specific sites reported in Figure 2 were all inferior to the sylvian fissure on the lateral surface of the superior temporal gyrus (S1: electrodes 23 and 14, 5.66 mm center-to-center; S2: electrodes 56 and 63, 5.66 mm center-to-center; S3:

electrodes 31 and 39, 4 mm center-to-center). During the recording of subject S2 the grid had to be repositioned by the surgeon, as a result the phoneme repetition task was recorded in a different position than the word repetition and passive phoneme task. The word-specific results for subject S2 compared passive phoneme listening to active word repetition. Word-specific sites were still found when comparing two nearby electrodes (< 1 cm) in the phoneme repetition and word repetition tasks (Wilcoxon rank-sum, $p<0.001$). Speech specific sites reported in Figure 4 were inferior to the sylvian fissure on the lateral surface of the temporal gyrus (S3: electrodes 40 and 48, 4 mm center-to-center; S4: electrodes 19 and 27, 4 mm center-to-center).



Supplemental Figure 4
Intraoperative image with (left) and without (middle) the electrode array in subject S1 alongside the reconstructed MRI and corresponding grid orientation (right).

Data Acquisition

Electrophysiological and peripheral auditory channels were acquired using a custom made Tucker Davis Technologies (http://www.tdt.com/) recording system (256 channel amplifier and Z-series digital signal processor board). Channels were sampled at 3051 Hz and the peripheral auditory channels were sampled at 24.4 KHz. Electrophysiological data was recorded using a subdural electrode as reference (corner grid electrode) and an additional electrode as ground.

Data Preprocessing

All EEG channels were manually inspected by a neurologist in order to identify epileptic channels as well as epochs of ictal activity which spread to other channels. Channels contaminated by sustained epileptic activity, electrical line noise (60 Hz) or abnormal signal were removed from further analysis. The raw time series, voltage histograms, and power spectra were used to identify noisy channels. All 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 non-epileptic channels were also removed from further analysis. Speaker and microphone channels recorded simultaneously with EEG activity were manually inspected in order to mark onset of stimulus and 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 were removed from analysis, similarly trials overlapping with ictal activity were discarded. All analyses were done

using custom scripts written for MATLAB, The Mathworks Inc. (http://www.mathworks.com/).

Data Analysis

Event Related Spectral Perturbations (ERSP) were created by first computing power series across the entire task for several spectral bands, producing a time-frequency representation of the entire task. Subsequently event related power averages locked to stimuli onset were computed and baseline corrected (-300 -> -100 ms) for each spectral band using a temporal window of -100 -> 750 ms in the passive tasks and -100 -> 2200 in the repetition task. Spectral bands were defined using center frequencies logarithmically spaced from 1 through 250 Hz and a fractional bandwidth of 20% the center frequency (i.e. a band centered at 4 Hz will be 3.6-4.4 Hz and a band centered at 100 Hz will be 90-110 Hz). Statistical significance was assessed using a bootstrapping approach similar to the method used by Canolty et al., 2007. In brief, an ERSP is created using N randomly generated time stamps (time windows do not overlap with ictal activity) instead of the N real stimuli onsets (N=number of real stimuli onsets). This process is then repeated over 1000 times producing a surrogate distribution of ERSPs. Each time-frequency point in the real ERSP can now be transformed using a z-score based on the mean and standard deviation of the surrogate distribution (which is normally distributed). All time-frequency plots are corrected for multiple comparisons (number of frequencies, time points and channels) using False Discovery Rate (FDR) correction of q=0.05 (Benjamini and Hochberg, 1995).

In order to assess statistical differences between ERSPs locked to phonemes and words the raw power values were directly tested using a non-parametric Wilcoxon rank sum test. For each stimuli onset an average was computed across a temporal window of 0->200 ms and a spectral window of 75.9 -> 144.5 Hz (all center frequency bands within range). The average power values for all phoneme stimuli were tested against those of all word stimuli. Furthermore, an analysis of variance was run on raw power values across a temporal window of 0->400 ms and a spectral window of 75.9 -> 144.5 Hz to assess condition X electrode interactions. Interactions were statistically assessed using a post-hoc pair-wise t-test.

The same approach was used in order to compare responses locked to hearing and speaking phonemes (a temporal window of 0->200 ms and a spectral window of 75.9 -> 144.5 Hz)

Onset and duration times of γ_{High} responses locked to phonemes and words were assessed during the first 750 ms post-stimulus using a spectral window of 75.9 -> 144.5 Hz averaged across trials.

Single trial γ_{High} 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 -300 -> -100 ms pre-stimulus). Statistical differences

between hearing and speaking were assessed by comparing average γ_{High} responses within a temporal window of 0-> 200 ms locked to stimuli.

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).

Results

Intraoperative electrocorticographic (ECoG) activity was recorded from four subjects undergoing neurosurgical treatment. Cortical responses were sampled from a high-density multi-electrode grid (inter-electrode spacing of 4 mm) placed over the posterior superior temporal gyrus (STG). Figure 1 depicts cortical responses indexed by oscillatory high gamma (γ_{High}: 70-150 Hz) activity across a 64 contact grid in subject S1 performing a passive listening task to phonemes (top) and a separate task listening to words (bottom). Significant γ_{High} activity commenced as early as 100 ms post stimulus onset and varied in temporal onset across different electrodes (z-scores corresponding to $p<0.001$ after FDR correction, see Methods). Cortical sites responding to phoneme stimuli showed a consistently similar spectral response pattern for word stimuli. γ_{High} activity locked to word stimuli remained active longer than γ_{High} activity locked to phoneme stimuli. Sustained γ_{High} activity locked to words remained active throughout stimulus presentation and in some sites up to 200 ms after stimulus has ended. Several cortical sites, which did not produce a significant response during phoneme listening, produced a robust response during word listening. Furthermore, many of these word specific sites were surrounded by electrodes 4 mm away that responded to both stimuli types.

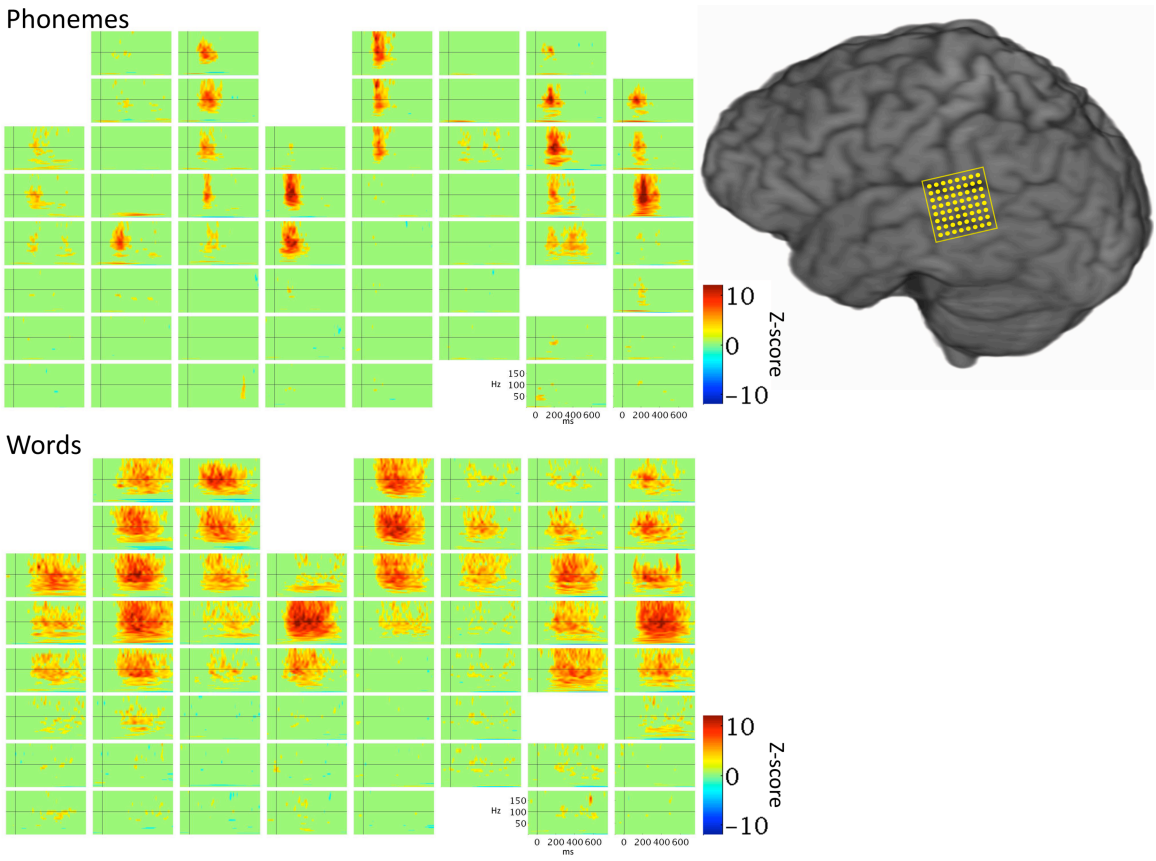


Figure 1
Spatiotemporal responses to phonemes (top) and words (bottoms) across a 64 contact 8x8 electrode grid in subject S1. Event related spectral perturbations are shown for each electrode locked to the onset of stimuli. Color scale represents statistically significant changes in power compared to a bootstrapped surrogate distribution (only significant data is shown, $p<0.05$ after FDR multiple comparison correction). Electrodes with no contact or abnormal signal are not shown.

All three subjects who performed both a phoneme and a word task (see Supplemental Table 1) exhibited cortical sites with word specific responses. While one site showed robust activity for both phoneme and word stimuli (Figure 2, Electrode B), an adjacent electrode revealed significant γ_{High} responses for words and minimal responses for phonemes (Figure 2, Electrode A). The word specific electrodes (Electrode A in each subject, see Supplemental Figure 1 for electrode anatomical positions) showed a significant difference between word and phoneme γ_{High} responses within a time window of the first 200 ms (Wilcoxon rank-sum, $p<0.001$). Furthermore, we ran an analysis of variance within each subject in order to assess a condition (phoneme or word) across electrode (A or B) interaction. We found a significant condition X electrode interaction within each subject (S1: $F(1,698) = 6.68$, $p<0.01$; S2: $F(1,598) = 126.76$, $p<0.001$; S3: $F(1,454) = 9.33$, $p<0.01$). For each subject a main effect of condition was significant (S1: $F(1,698) = 22.7$, $p<0.001$; S2: $F(598) = 30.2$, $p<0.001$; S3: $F(454) = 20.99$, $p<0.001$) as well as a main effect of electrode (S1: $F(1,698) = 45.99$, $p<0.001$; S2: $F(598) = 32.27$, $p<0.001$; S3: $F(454) = 61.61$, $p<0.001$). A breakdown of the interactions revealed that all subjects showed a significant difference between electrodes in both the phoneme and word conditions as well as a significant difference between conditions within electrode A

(a detailed breakdown of the interactions and the main effects are provided in Supplemental Figure 2).

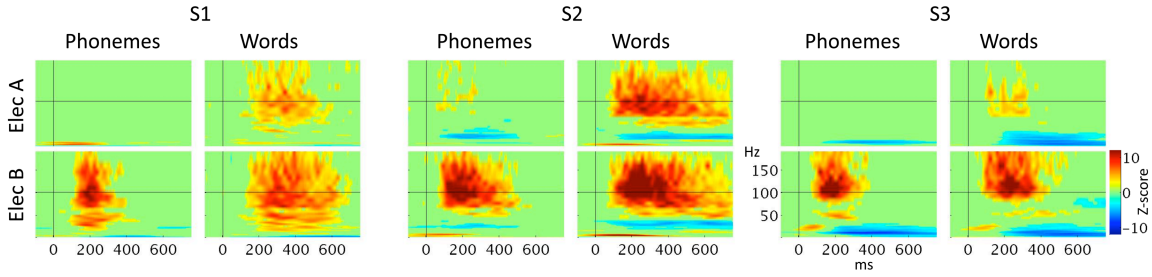
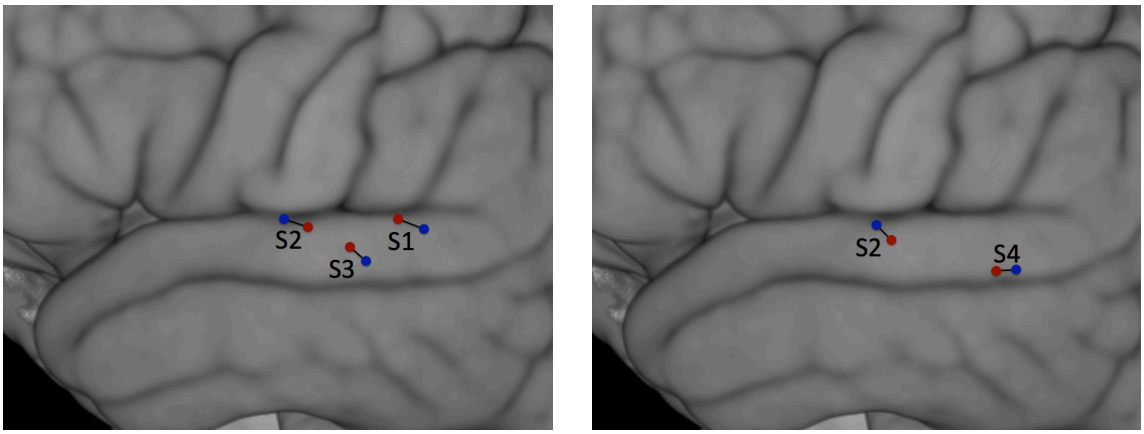


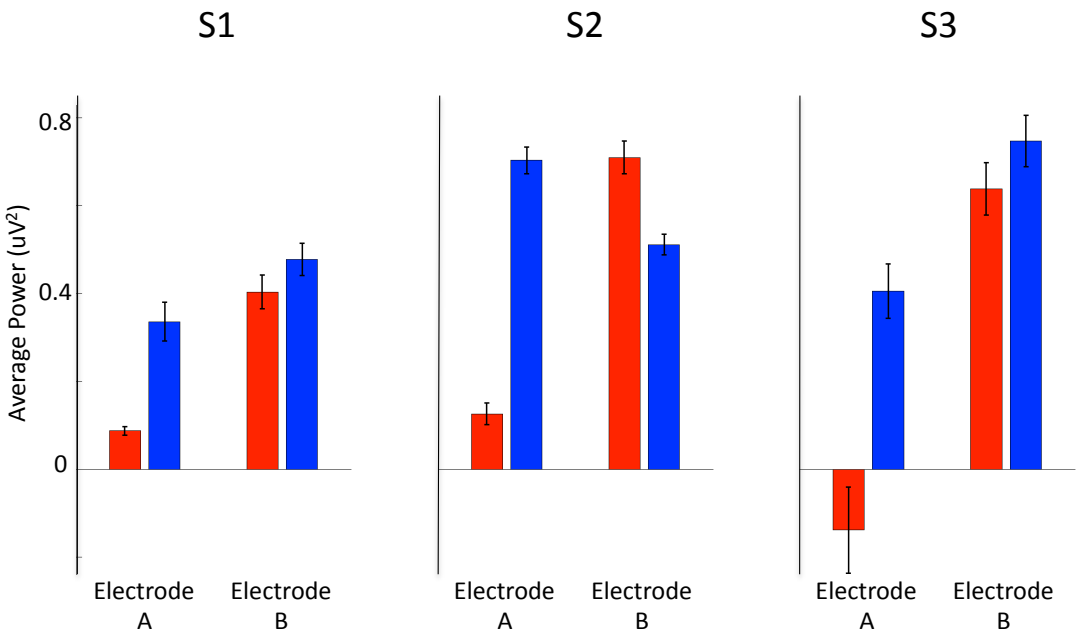
Figure 2

Event related spectral perturbations of two adjacent electrodes in three subjects locked to phonemes and words. Electrode A (top row) responds selectively to word stimuli while Electrode B (bottom row) 4 mm away responds to both stimuli types. Vertical line marks stimuli onset, horizontal line marks 100 Hz, the color scale represents statistical significant changes in power compared to a bootstrapped surrogate distribution. Only statistically significant values are shown ($p<0.05$, after FDR multiple comparison correction). See Supplemental Figure 1 for electrode anatomical positions.



Supplemental Figure 1

Left: Anatomical positions of electrode pairs reported in Figure 2; Red marks electrode A (word-specific site) and blue marks electrode B. Right: Anatomical positions of electrode pairs reported in Figure 4; Red marks electrode A (self-speech site) and blue marks electrode B.



Supplemental Figure 2
Average power estimates for each electrode and condition (phonemes marked in red and words marked in blue). For each subject a main effect of condition was significant (S1: $F(1,698) = 22.7$, $p<0.001$; S2: $F(598) = 30.2$, $p<0.001$; S3: $F(454) = 20.99$, $p<0.001$). For each subject a main effect of electrode was significant (S1: $F(1,698) = 45.99$, $p<0.001$; S2: $F(598) = 32.27$, $p<0.001$; S3: $F(454) = 61.61$, $p<0.001$). A breakdown of the interaction revealed a significant difference between electrodes in the word condition (S1: $t(298) = -2.46$, $p<0.05$; S2: $t(198) = 4.97$, $p<0.001$; S3: $t(228) = -4.03$, $p<0.001$) and in the phoneme condition (S1: $t(400) = -8.00$, $p<0.001$; S2: $t(400) = -13.08$, $p<0.001$; S3: $t(226) = -6.77$, $p<0.001$). A significant difference between conditions was found within electrode A (S1: $t(349) = -6.22$, $p<0.001$; S2: $t(299) = -14.12$, $p<0.001$; S3: $t(227) = -4.70$, $p<0.001$). Subjects S1 and S3 did not show a significant difference between conditions in electrode B while subject S2 did show a significant difference (S1: $t(349) = -1.35$, $p=0.18$; S2: $t(299) = 3.58$, $p<0.001$; S3: $t(227) = -1.31$, $p=0.19$).

Three of the four subjects performed a phoneme repetition task (see Supplemental Table 1). Figure 3 depicts cortical responses evoked by hearing and then producing phonemes across a 64 contact grid in subject S4. A majority of electrodes exhibited a γ_{High} response locked to hearing phonemes commencing as early as 80 ms post stimulus onset. Responses locked to phoneme production varied across electrodes. A majority of cortical sites exhibited a suppressed response during production as compared with hearing (Wilcoxon rank-sum, $p<0.001$), while several sites did not show a significant change in response (Wilcoxon rank-sum, $p>0.05$). One electrode exhibited an enhanced response during production while its eight neighboring electrodes showed a suppressed response as compared with hearing (Figure 3, blue box; arrowhead). A direct comparison of spectral γ_{High} responses locked to hearing and speaking is detailed in Supplemental Figure 3.

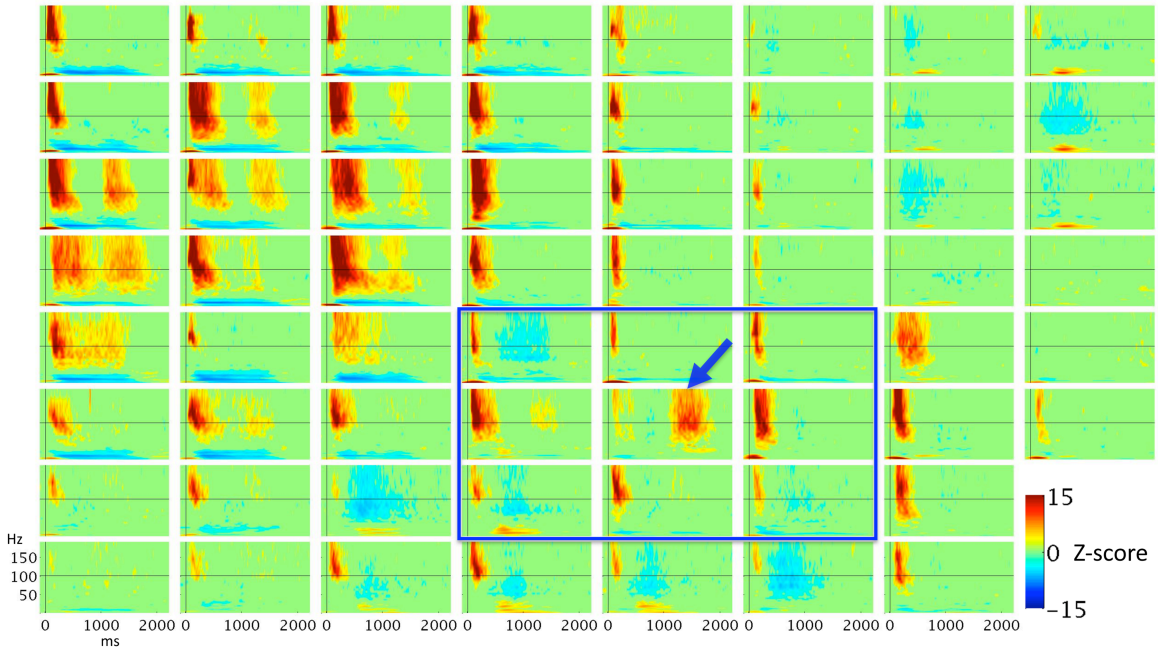
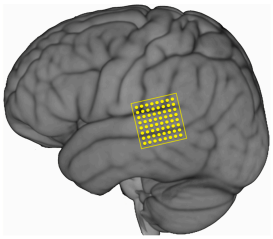
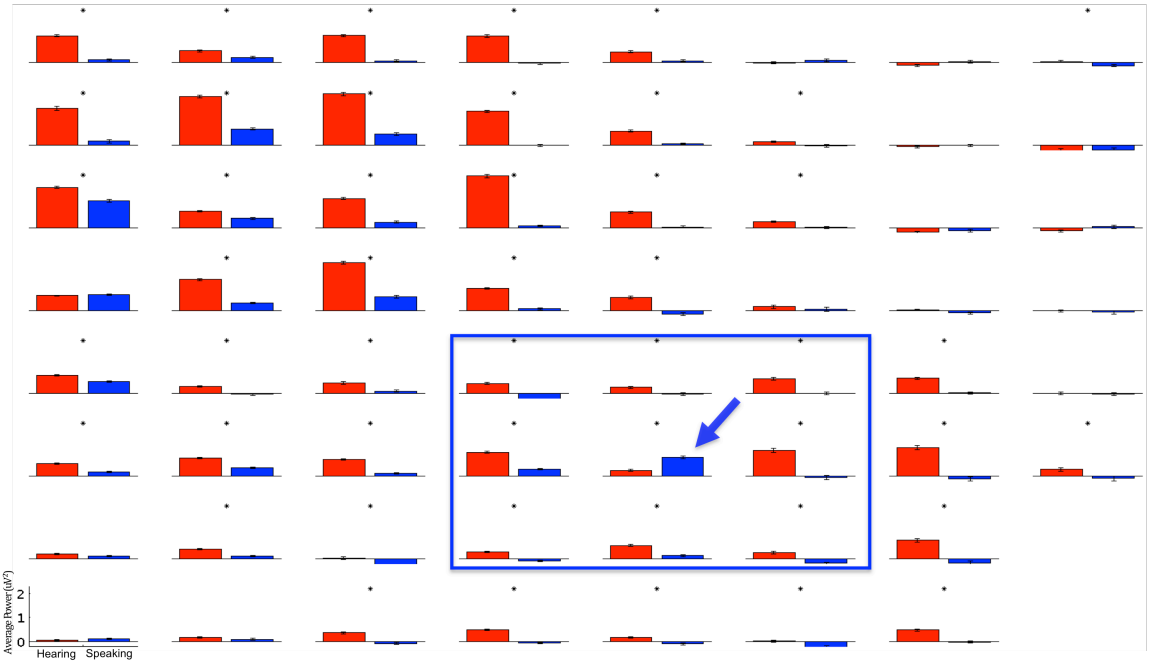


Figure 3.



Spatiotemporal responses to hearing and then producing phonemes across a 64 contact 8x8 electrode grid in subject S4. Event related spectral perturbations are shown for each electrode locked to the onset of phonemes. Color scale represents statistically significant changes in power compared to a bootstrapped surrogate distribution (only significant data is shown, $p<0.05$ after FDR correction). Blue box marks an area of approximately 1 cm^2 of cortex; blue arrowhead marks an electrode exhibiting self-speech specificity. Electrodes with no contact or abnormal signal are not shown.



Supplemental Figure 3

Average spectral γ_{High} estimates for each electrode locked to hearing (red) and speaking (blue) phonemes across a 64 contact 8x8 electrode grid in subject S4 (identical grid orientation depicted in Figure 3). Stars

mark electrodes with a significant difference between hearing and speaking conditions (Wilcoxon rank-sum, $p<0.001$). Blue box marks an area of approximately 1 cm^2 of cortex; blue arrowhead marks an electrode exhibiting self-speech specificity. Electrodes with no contact or abnormal signal are not shown.

Two subjects revealed cortical sites that exhibited production specific responses while a neighboring electrode produced an opposite functional response (see Supplemental Figure 1 for electrode anatomical positions). Single trial analysis of the responses showed consistent responses across trials as seen in Figure 4. Stacked γ_{High} power traces for each trial locked to hearing phonemes are shown for two adjacent electrodes in subject S4 (top) and S2 (bottom). A production specific electrode (electrode A) exhibited significantly higher γ_{High} power after production (black line) as compared with hearing external stimuli (Wilcoxon rank-sum, $p<0.001$). The adjacent electrode (electrode B) at 4 mm distance exhibited a robust response to the external stimuli (relative to baseline) while the response to self-produced speech is severely reduced (relative to hearing in both subjects and relative to baseline in subject S4).

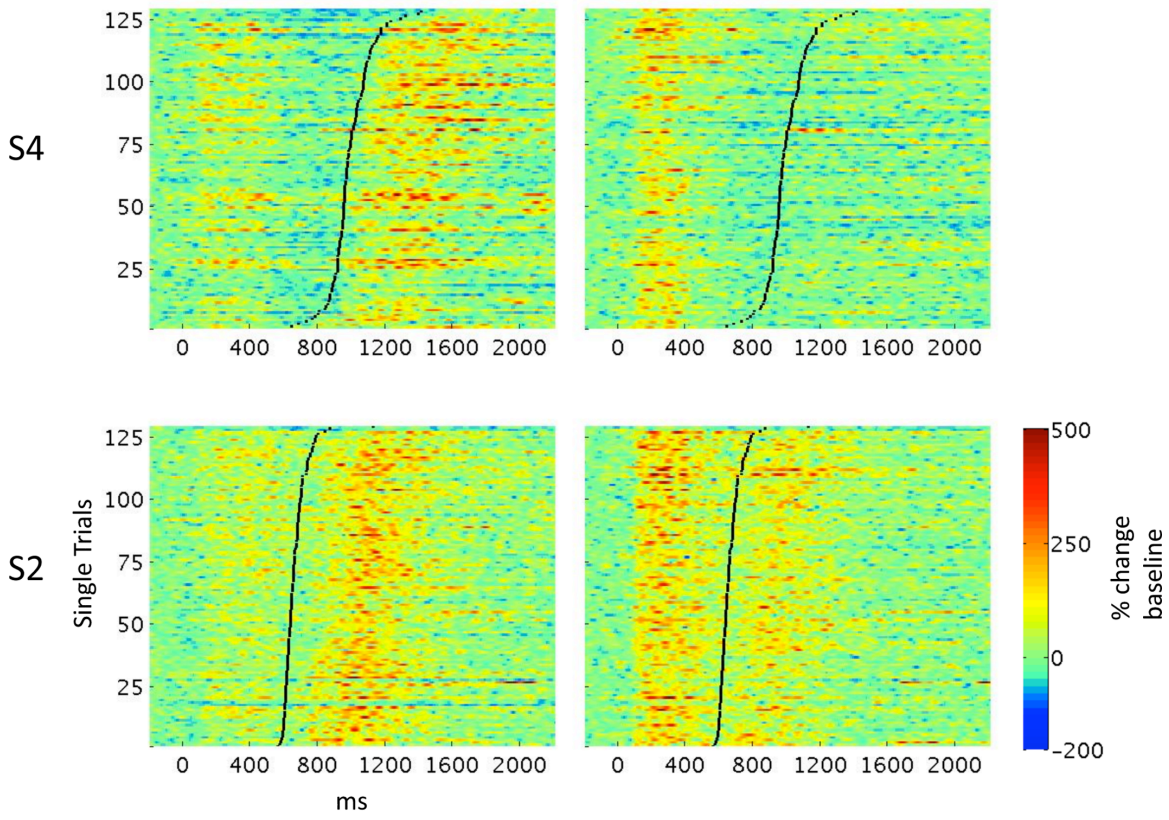


Figure 4

Single trial γ_{High} power traces vertically stacked for two adjacent electrodes in subjects S2 and S4. Zero ms marks phoneme onset and black lines mark production onset (trials are sorted for display purposes). Color-scale represents percent change from the pre-stimulus baseline in each trial. See Supplemental Figure 1 for electrode anatomical positions.

Discussion

We recorded electrocorticographic (ECoG) activity directly from the surface of the temporal cortex of four awake patients undergoing neurosurgery. The aim of the study was to assess the temporal-spatial distribution of cortical activity during processing of phonemes and words. Recording ECoG activity from high-density electrode grids revealed different functional responses of cortical sites separated by 4 mm distance. All three subjects who performed both a word and phoneme related task showed word specific activity adjacent to cortical sites that were non-specific to stimulus type. Two of the three subjects who performed a phoneme repetition task showed production specific activity adjacent to cortical sites that were active during hearing and suppressed during production. These functionally distinct responses within 4 mm of cortex support a dense distribution of independent sub-regions supporting linguistic processing. Evidence of such dense distribution of activity supports a spatially fine graded cortical representation of speech processing in the human temporal lobe. Neuroanatomical differences across subjects in this fine graded representation may obscure current neuroimaging findings, which typically employ averaging techniques over many subjects.

The superior temporal gyrus (STG) and superior temporal sulcus (STS) have long been implicated in both word processing as well as phonological processing (Wise et al., 1991; Démonet et al., 1992; Zatorre et al., 1992; Boatman et al., 1995; Price et al., 1996; Binder et al., 1997). Studies in non-human primates have shown that superior temporal neurons are broadly tuned and exhibit stronger responses to wideband noise than pure tones. Furthermore, neurons in the lateral surface of the STG are selective to species-specific vocalization (Rauschecker et al., 1995; Rauschecker, 1998). Findings in humans have shown superior temporal lobe specificity to intelligible speech (Mummery et al., 1999; Binder et al., 2000; Scott et al., 2000; Canolty et al., 2007). Our ECoG data, which is sampled from neuronal populations in the lateral surface of the STG, provides similar evidence for complex stimulus dependent local specificity. While one neuronal population responds similarly to both phoneme and word stimuli a neighboring population responds selectively to words. This result can be explained by the attributes of the stimulus presented as well as the specificity of the underlying neuronal populations. One hypothesis is that the neurons in the two neighboring populations have different spectral tuning. The word selective responses could be the result of a large neuronal population responding to wider and more specific frequency bands in the stimuli. An alternate hypothesis is that one population contains more neurons which are responsive to a specific class of complex stimuli. These two hypotheses are not mutually exclusive and could both contribute to the observed neurophysiological signals. Furthermore, we cannot rule out the possibility that the different interstimulus intervals (ISI) in the phoneme (1.5 s ISI) and word (2 s ISI) tasks may partially contribute to the differential functional response (Ceponiene et al., 1998; Röder et al., 1999). Our results, which complement previous findings in non-human primates (Rauschecker et al., 1995; Rauschecker, 1998), provide further evidence of a hierarchical organization of the temporal cortex where different neuronal populations preferentially respond to more complex stimuli.

The human temporal cortex has consistently shown a suppressed pattern of activity during self-produced speech as compared with external stimuli (Curio et al., 2000; Ford et al., 2001; Houde et al., 2002; Christoffels et al., 2007). Single unit studies in both non-human and human primates have provided evidence of STG neurons that are suppressed during vocalization as well as a subset of neurons which do not show suppression (Müller-Preuss 1981; Eliades 2003; Eliades 2008; Creutzfeldt 1989). Our results showing suppressed auditory responses during production replicate previous ECoG findings (Crone 2001; Towle 2008) as well as non-invasive electrophysiological results (Ford 2001; Numminen 1999; Houde 2002). The fact that we find cortical sites exhibiting both suppressed, non-suppressed and excited responses could be due to variability in the proportion of suppressed neurons in the sampled region of cortex. Such an interpretation could explain the differing proportions of suppressed neurons reported in single unit studies (Müller-Preuss and Ploog, 1981; Creutzfeldt et al., 1989; Eliades and Wang, 2003; Eliades and Wang, 2008). The speech-specific cortical sites we have found in two patients provide evidence for monitoring of self-speech. These electrophysiological responses could represent a population of neurons with specificity to self-produced vocalization. We currently cannot rule out the possibility that these responses are due to neuronal tuning to the spectral attributes of self-speech such as the acoustic masking of the auditory signal in the middle ear due to bone conduction (v. Békésy, 1949). Nevertheless, our results provide further evidence for a rich dynamic of spatial cortical responses in the STG, which are functionally distinct.

Evidence from both non-human primates (Hackett et al., 1998; Rauschecker, 1998; Romanski et al., 1999) as well as humans (Scott et al., 2000; Wise et al., 2001; Hickok and Poeppel, 2007) supports a hierarchal organization of functionally distinct subdivisions of the auditory cortex. The current emphasis is on anterior and posterior functional streams of processing similar to those in the visual domain (Rauschecker and Tian, 2000; Hickok and Poeppel, 2007). Our current data supports a richer spatial organization within the temporal cortex suggesting a more complex and finer functional subdivision.

Chapter 4

Redefining the role of Broca's area in speech production

Abstract

For many decades, the neurobiological basis of language has been dominated by a conceptually dichotomous model in which speech perception is supported by Wernicke's area in the temporal lobe and speech production is supported by Broca's area in the frontal lobe (Broca, 1861; Wernicke, 1874; Geschwind, 1970). More recently, this model has been challenged by lesion and neuroimaging studies suggesting a more complex network of cortical structures supporting language and implicating Broca's area in both receptive and expressive function (Friederici 2002, Hickok and Poeppel 2007). Understanding how Broca's area contributes to these functions critically depends on the temporal dynamics of its recruitment, which have remained largely unknown. Using cortical surface recordings in neurosurgical patients, we elucidate the timing and magnitude of neural activity in Broca's area relative to both speech perception and production. We show that Broca's area is not active during articulation but is robustly active during an intermediate stage that overlaps perception but precedes articulation. These unique electrophysiological data suggest that Broca's area may serve as an interface between perception and production, supporting articulatory preparation but not involved in the actual act of speaking, providing a new view of language processing in the human cortex.

Introduction

The seminal findings of Paul Broca and Carl Wernicke over a century ago provided critical insights into the neurobiological basis of human language. Their observations heavily influenced current neuropsychological models of language processing, assigning perception of speech to posterior temporal cortex (Wernicke's area; superior temporal gyrus, superior temporal sulcus) and speech production to inferior frontal cortex (Broca's area; pars opercularis and pars triangularis). However, converging evidence from lesion, neuroimaging and neurosurgical studies have questioned this dichotomy, particularly the exclusive role of Broca's area in speech production. Lesion studies have shown that cortical damage limited to Broca's area does not cause a Broca's aphasia but rather results in a transient, rapidly improving mutism (Mohr et al., 1978), and that articulation deficits do not necessarily involve damage to Broca's area (Dronkers, 1996). Likewise, intraoperative electrical stimulation mapping at sites outside Broca's area can interfere with speech output (Ojemann et al, 1989), and stimulation of Broca's area itself can elicit errors in perception with preserved speech output (Sanai et al. 2008). Moreover, PET and fMRI studies have shown consistent Broca's area activation in receptive tasks that do not require overt articulation (Zatorre et al., 1992; Price et al., 1996; Binder et al., 1997). Elucidating the cognitive functions Broca's area supports during perception, articulatory preparation and speech production critically depends on its temporal dynamics. To date, the timing of activity in Broca's area has been mostly derived from neuroimaging and non-invasive electrophysiological studies. While these studies have provided valuable estimates as to when Broca's becomes active during language tasks (Friederici, 2002; Indefery and Levelt, 2004), they have not clarified the role of Broca's area in speech. A combined temporal and spatial resolution is necessary for determining the components of speech production supported by Broca's area, as well as how these relate to speech perception.

Electrocorticographic (ECoG) recordings acquired from the surface of the brain offer precisely that opportunity. During a unique neurosurgical procedure, electrodes are implanted directly on the surface of the cortex in order to surgically manage patients with medically-refractory epilepsy. These electrodes provide an invaluable opportunity to record neural activity with unprecedented spatial and temporal resolution. The exceptional signal quality afforded by these recordings also allow for single-trial analyses of cortical population activity, which is ideal for investigating the spatio-temporal dynamics of language. A few ECoG studies have reported frontal lobe activity during speech production (Fried et al., 1981; Towle et al., 2008; Edwards et al., 2010; Pei et al., 2011). While these studies have provided evidence for frontal involvement, the specific temporal window of cortical recruitment for Broca's area has not been defined. A recent intracranial study (Sahin et al., 2009) was the first to focus on Broca's area (Llorens et al., 2011) revealing its involvement in lexical, grammatical and phonological processing (without articulation) in an early post-stimulus temporal window (200-400 ms). Nevertheless, the study did not address the role of Broca's area in speech production. We address this issue using electrocorticographic (ECoG) recordings obtained directly from the surface of the cortex providing a robust neurophysiological signal for analysis of the brain dynamics underlying speech (Flinker et al., 2010). Seven patients with electrode implantations over peri-sylvian language regions including Broca's area consented and

participated in the study during lulls in clinical treatment. Subjects participated in a battery of tasks involving perception as well as overt articulation of words, including repetition and reading aloud (see Methods and Figure 1 for details).

Methods

Subjects Tasks and Stimuli

Seven subjects (S1-7) undergoing neurosurgical treatment for refractory epilepsy at Johns Hopkins Hospital participated in the study (see Supplemental Table 1). Electrode placement and medical treatment were dictated solely by the clinical needs of each patient. All subjects were left-hemisphere dominant for language and had left-hemisphere electrode coverage. Subjects S1-3 and S5-6 took part in a previously described word repetition task (Flinker et al., 2010, 2011), which consisted of repeating aurally presented words, three phonemes in length (mean duration=525ms, SD=100ms). Subjects S1-5 and S7 took part in a previously-described word reading task (Flinker et al., 2010, Crone et al., 2001b), which consisted of reading aloud visually-presented mono- and bi-syllabic words. Subjects S2, S3, and S5 participated in an additional word repetition task, which consisted of words varying in length (3-10 phonemes). Finally, Subjects S1-3 and S6 participated in a target detection task, which included the same stimuli as in the first word repetition task; the subjects were asked to listen to all words and press a button when they heard a target word (one target appearing at 0.08 probability; target stimuli were removed from analysis). All peripheral signals and responses were recorded together with intracranial EEG signals to ensure proper synchronization (sampled at 1000 Hz using a clinical 128-channel Harmonie system; Stellate, Montreal, Canada).

Supplemental Table 1

Subject	Age	Sex	Handedness
S1	18	Male	Right
S2	38	Male	Left
S3	18	Male	Right
S4	14	Male	Right
S5	36	Female	Right
S6	28	Female	Right
S7	29	Female	Right

Electrode Localization

A structural preoperative MRI as well as a post-implantation CT scan was acquired for all subjects. The MR and CT scans were co-registered to the same space using two non-linear transforms based on normalized mutual information implemented in the Bioimage suite (Papadetris et al.) (the second transform was used to correct for slight shifts in brain morphology caused by the electrodes). The results were then compared to an intraoperative photo image of the exposed grid after it was sutured to the dura. Electrodes were classified according to anatomical location (superior temporal gyrus,

precentral gyrus, pars opercularis, pars triangularis, middle frontal gyrus) within each subject's anatomical space.

Data Analysis

All electrodes containing sustained ictal activity as well as trials with transient epileptiform activity or no behavioral response were removed. All remaining electrodes covering the frontal cortex and the superior temporal gyrus were included in the analysis. Electrodes were defined as significant if they showed a statistical difference (two-sample t-test, α=0.001, Bonferroni corrected, see reference Flinker et al., 2010) in at least one of seven 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) by comparing log-transformed power during pre-stimulus baseline (-450ms -> -50 ms) with the post-stimulus epoch (0 ms -> speech response + 500 ms). Power spectral densities were computed for the baseline and post-stimulus epochs to assess event-related changes in the frequency domain (Figure S1). Event-related spectral perturbations were computed using log-transformed power (Figure 1) as previously reported (Flinker et al., 2011) and were assessed for significance using a bootstrapping approach comparing power estimates to pooled distributions from baseline (see Statistical Bootstrapping). Based on our results showing peak activity at 100 Hz (Figures 1,S1) as well as previously reported findings (Towle et al., 2008; Edwards et al., 2010; Flinker et al., 2010, 2011; Pei et al., 2011) we focused on the γ_{High} band (γ_{High}: 70-150 Hz). Averaged event-related log-transformed γ_{High} traces were computed (Hilbert transform, see reference Flinker et al., 2010), smoothed using a Hanning window (100 samples), and transformed to units of z-score significance compared to a bootstrapped baseline distribution (see Statistical Bootstrapping). A peak γ_{High} value was defined as the maximum value within a significant window (a minimum of 100 ms of contiguous points passing a significance threshold corresponding to α=0.0023) and was calculated separately locked to stimulus onset and speech production. Onsets and offsets of γ_{High} activity were computed by taking the first and last time sample that passed significance.

Statistical Bootstrapping

For each subject, averaged power estimates for all trials within a specific task and electrode were compared to a bootstrapped distribution of pre-stimulus baseline power values within each frequency band (-250 ms -> -50 ms). N random samples were pooled from all the baselines and averaged to produce a surrogate power sample (where N is the number of trials within a specific task). This process was repeated 1000 times to create a surrogate distribution with a normal distribution. Real power estimates (post-stimulus) were compared to this distribution to assess significance. For the event-related spectral perturbations, all time and frequency significance values were corrected for multiple comparisons using an FDR correction (q=0.05, Benjamini and Hochberg, 1995). Averaged event-related γ_{High} traces were computed in the same manner without FDR correction; instead, a threshold for contiguous significant samples was used (100 samples with a p-value of 0.0023).

Single Trial analysis

Single trial γ_{High} traces were computed for all electrodes with a peak γ_{High} value in every subject and task (STG electrodes were excluded for the visual reading task). The log transform of the γ_{High} power time series was smoothed using a Hanning window (100 samples) and transformed to units of z-score compared to a pooled baseline (-250 ms -> -50 ms) distribution of all trials within that block. This transforms single trial samples to units of significant activity within that single trial (compared to the normal distribution formed by all baseline samples). In order to assess the temporal lag between anatomical sites we computed cross-correlations for each single trial between electrode pairs (within subject and task) that were then averaged to produce an averaged cross-correlation for that electrode pair. The time point of maximum cross-correlation value was defined as the temporal lag for that electrode pair. Only temporal lags for a significant cross-correlation were included in the final results. Significance was assessed using a permutation approach. For each electrode pair, the cross-correlation was calculated using single trials that were randomly shuffled between the two electrodes. This process was repeated 1000 times to produce a distribution of random cross-correlation values that were used to assess non-parametric significance of the real correlation values.

Results

We first examined robust increases in high frequency power (γ_{High}: 70-150 Hz; γ_{High} provided the most reliable spectral measure of cortical activation; see Methods and Figures 1,S1) across the peri-sylvian language cortex in an auditory word repetition task. Cortical activity exhibited a systematic temporal propagation from auditory cortices (superior temporal gyrus; STG and superior temporal sulcus; STS) to Broca's area (pars triangularis and opercularis), eventually reaching premotor and motor cortex during word articulation. Figure 1 presents this pattern in a representative subject during perception and subsequent production of words. Auditory cortex activity (indexed by γ_{High}) commenced as early as 65 ms after stimulus onset and peaked within 200 ms. Activity in Broca's area closely followed auditory activation, starting within 240 ms of stimulus onset and peaking at 340 ms, during which time the spoken word stimuli were still being presented. Activity in Broca's area ended prior to speech onset. Motor cortex activity, in contrast, was spread across the articulatory onset window (Mean RT = 1200ms) and was apparent prior to and during speech production. Importantly, Broca's area was active only during the time window between perception and production, and its neural activity was extinguished by the time of actual speech articulation (Figure 1a,b).

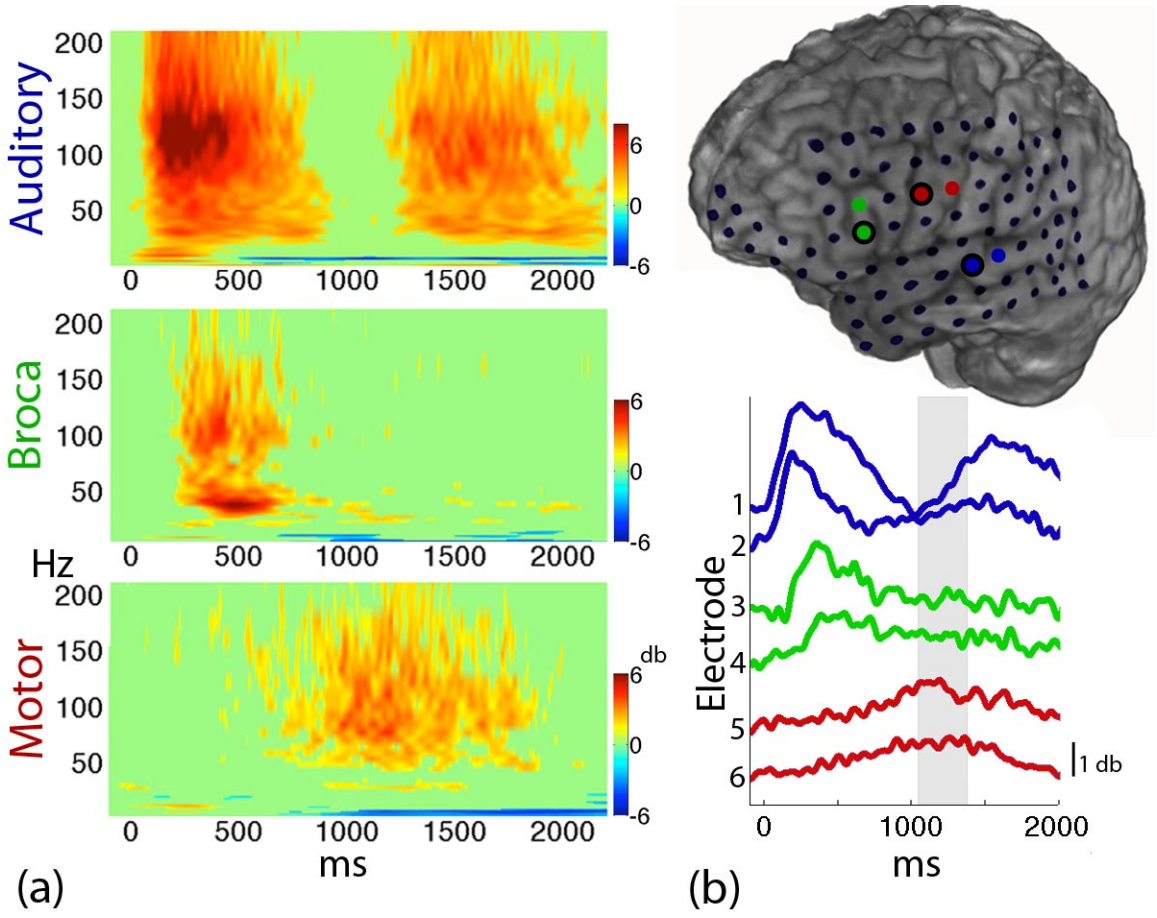


Figure 1

(a) Event-related spectral perturbations (ERSPs) locked to hearing word stimuli. Cortical activity indexed by power in high frequencies propagates from auditory cortex during word perception to Broca's area and extends to motor cortex during word production. (b) Averaged across trials, high frequency power (γ_{High}, 70-150 Hz) traces locked to hearing word stimuli are shown for auditory (blue), Broca (green) and motor (red) electrodes. The first electrode in every pair is marked by a black circle and corresponds to the ERSP plotted on the left. The shaded grey area marks articulation onset distribution (one standard deviation in each direction).

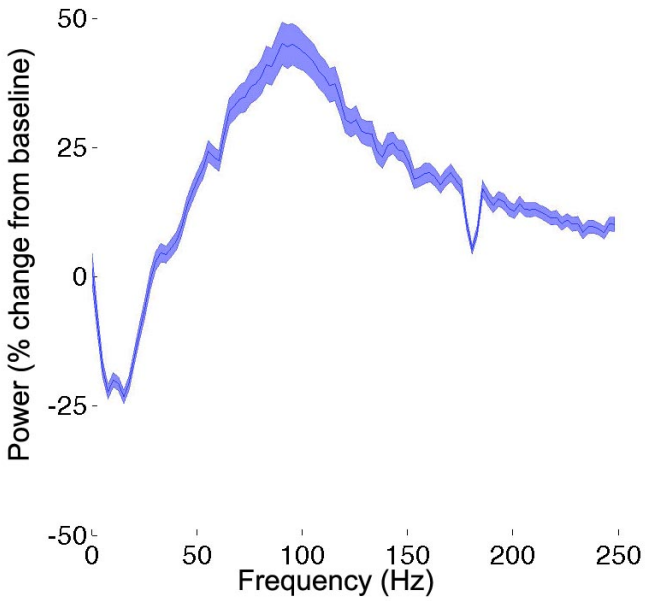


Figure S1

Average PSDs (power spectral densities) were computed for each electrode and task across all subjects. The PSDs were computed for event-related windows following stimulus presentation up to 400 ms after speech production (consecutive windows of 400 ms using the Welch method). A separate PSD was computed for the baseline epochs (400 ms prior to stimulus presentation). The event-related PSDs were transformed to units of percent change by subtracting and then dividing by the power estimates of the baseline PSD. The trace shows the mean PSD and SEM for all electrodes, tasks and subject (N=274).

To assess the relationship of activity within this early time window (200-600 ms; range of significant activity) with speech production, we examined neural activity time-locked to articulation onset across all production tasks (word repetition and reading). Significant electrodes (defined by an increase in power) across tasks and subjects were classified according to peak γ_{High} activation. Figure 2a depicts the spatial distribution of electrodes at which maximum γ_{High} power was evident at least 100 ms prior to articulation onset (blue) and those at which γ_{High} power peaked during and after articulation started (red). Electrodes covering Broca's area were active prior to the onset of articulation but were inactive by the time speech commenced. In order to quantify the temporal propagation across different cortical sites, all electrodes were also classified according to within-subject gyral anatomy (superior temporal gyrus and superior temporal sulcus; pars opercularis and triangularis and precentral cortex). The propagation of peak activity across these anatomical regions is plotted in relation to stimulus onset as well as articulatory onset in Figure 2b. A two-way analysis of variance (Perception and production activity X anatomy) confirmed that the latencies of peak activity were different for each anatomical category; $F(2,98)=137.1$, $p<0.001$. These results support cortical activity propagating from the STG to Broca's area and terminating in motor cortex.

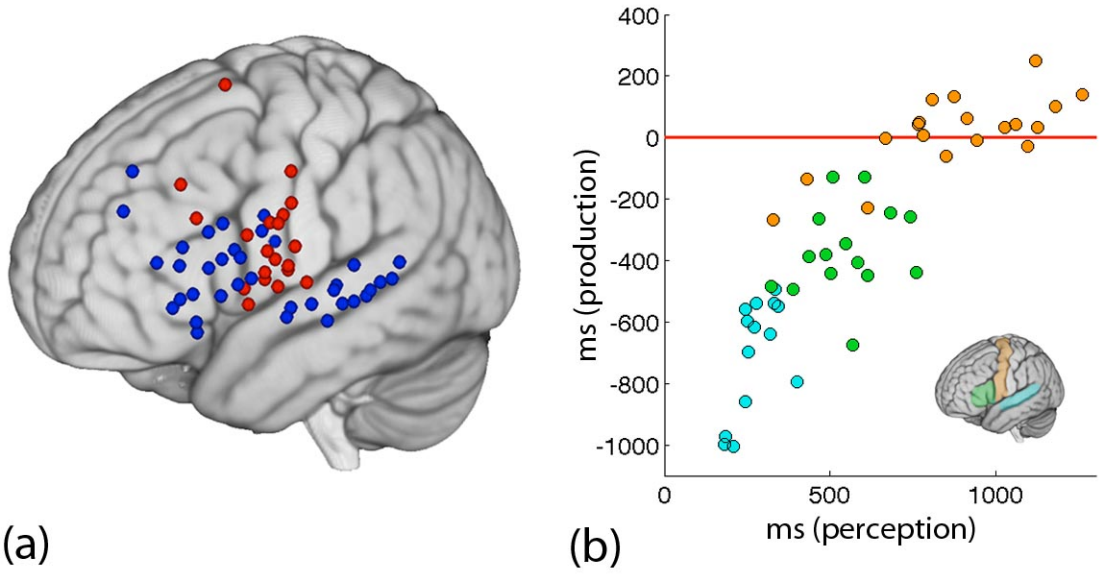


Figure 2
(a) The spatial distribution of temporal activity is shown for all subjects. Electrodes marked in blue peaked in activity prior to articulation onset while electrodes marked in red peaked in activity during and after articulation onset. (b) Peak activity locked to stimulus onset (x axis) and locked to speech onset (y axis) is displayed for electrodes in three anatomical locations: STG (cyan), Broca's area (green) and precentral gyrus (orange). Activity in both axes temporally propagates from STG to Broca's area and culminates in the precentral gyrus.

In order to assess the consistency of this pattern of averaged cortical responses, we investigated activation of Broca's area at the single-trial level. Individual single trials across all production tasks and subjects were grouped by anatomy and sorted according to response latency. In Figure 3, individual cortical responses from all trials are vertically-stacked, depicting a robust temporal pattern from perception to production on the single-trial level. We quantified the temporal lag between different cortical sites using the cross-correlation of single trials in each electrode pair (see Methods). The temporal lag between the STG and Broca's area was tightly distributed around 150 ms ($\mu = 149.6$ $\sigma=80.9$ SEM= 11.2). In contrast, the temporal lag between Broca's area and motor cortex was distributed around 350 ms, with a much larger variance ($\mu = 346.9$ $\sigma=287.2$ SEM= 47.2). The temporal lag between STG and Broca's area was different from the temporal lag between Broca's and motor areas (two sample t-test, $p<0.01$). These results support a temporal coupling between the superior temporal gyrus and Broca's area during language perception, and a more variable temporal coupling between Broca's area and motor cortices prior to speech production.

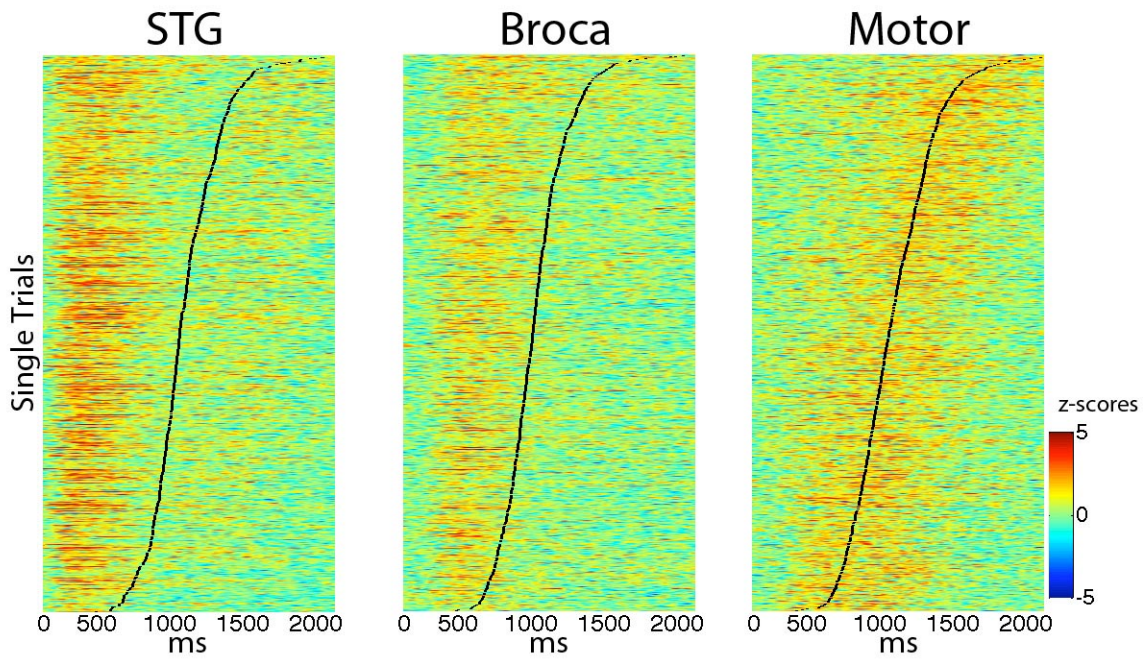


Figure 3

Vertically-stacked single trials are shown for all subjects and production tasks, sorted by response time (black line). Single-trial activity (z-scores within each trial compared to a baseline distribution) propagates from STG to Broca's area and culminates in motor cortex.

A critical question remained – was this early activity in Broca's area related to perception of stimuli or to early stages of production? We addressed this question in a subset of four subjects who participated in an auditory listening task (target detection without articulation) in addition to a separate repetition task of the same word stimuli. Activity in Broca's area was evident during both tasks. While activity in the repetition task requiring articulation was higher, substantial activity was still evident in the auditory listening task that did not require overt articulation (Figure 4a). Significant activity began 245 ms after stimulus onset in both tasks but lasted longer during the repetition task (580 ms of sustained significance; $p<0.05$, FDR corrected) than the listening task (280 ms). We were also able to compare this activity to a reading task performed by two of the subjects. Activity in Broca's area during auditory word repetition and visual word reading had similar temporal dynamics (Figure 4b,c) suggesting modality-independent processing in Broca's area. These dynamics are consistent with a role for Broca's area in both perceptual processing of linguistic information and the preparation of speech.

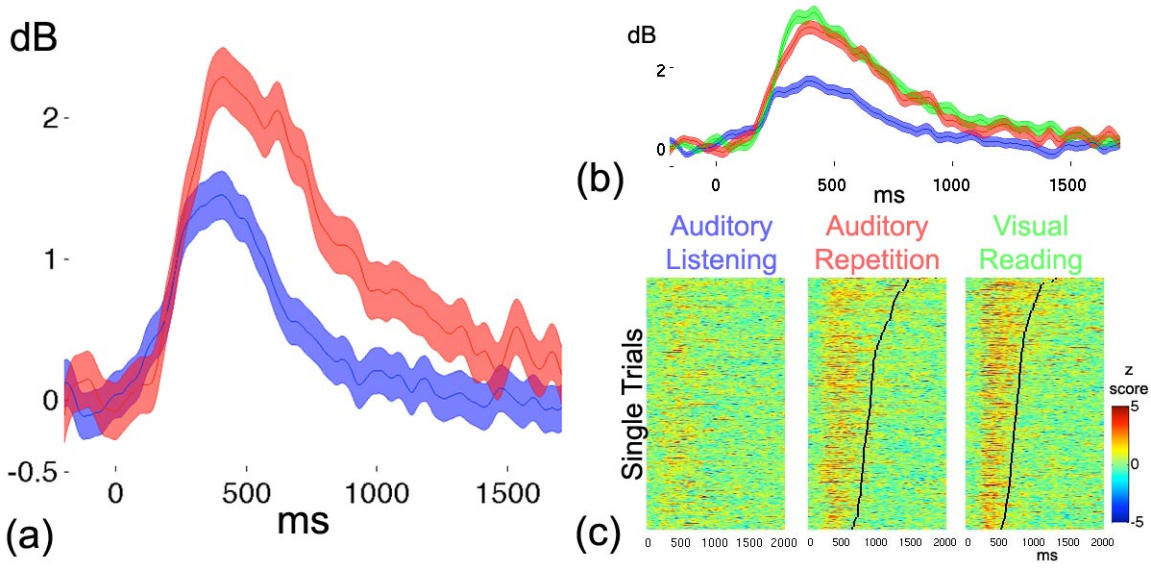


Figure 4

(a) Averaged cortical activity (γ_{High}) in Broca's area locked to hearing word stimuli in a repetition task (red trace) compared with a listening task that does not require articulation (blue trace). Shaded area denotes standard error of the means.

(b) Averaged cortical activity (γ_{High}) in two subjects locked to stimulus onset in an auditory word repetition task (red trace), a visual word reading task (green trace), and a listening task that does not require articulation (blue trace). Shaded area denotes standard error means.

(c) Vertically-stacked single trials are shown for the two subjects in Broca's area sorted by response time (black line). Single trial activity is in units of z-scores compared to a baseline distribution.

Discussion

Leveraging a unique temporal and spatial resolution of human neural population responses, our results implicate Broca's area in speech preparation but not in the actual act of articulation, contradicting the classical model of language organization in the human brain. The robust activity seen prior to articulation likely represents articulatory preparation and is consistent with recent models for speech production predicting a role for Broca's area in syllabification (Levelt et al., 1999;Indefrey and Levelt, 2004). Moreover, our findings across tasks reveal consistent activity during a novel intermediate stage that overlaps perception and precedes articulation. Such a stage is critical for facilitating transformations between auditory or visual perceptual representations and an articulatory code for speech production. This intermediate activity may not be limited to articulatory transformations and likely supports processes previously linked to Broca's area, including phonological segmentation, syntactic processing, and unification that involves integrating different types of linguistic information into one representation (Burton et al., 2000; Friederici, 2002; Hagoort, 2005). We propose that language is supported by a fronto-temporal network, with Broca's area acting as a mediator between perception and production rather than as the seat of articulation.

Chapter 5

Concluding Remarks

The neurobiological basis of language is an age-old mystery. Scientific and technological advances over the past century have driven numerous studies aimed at elucidating the functional neuroanatomy of language processing in the human cortex. Most experimental techniques available in humans are strictly non-invasive, constraining results in either the temporal or spatial domain. Acquiring a combined temporal resolution and accurate source localization requires invasive recordings, possible only during necessary clinical treatment in humans. While these recordings have their own limitations, they offer a golden opportunity to address novel questions and complement non-invasive findings.

The first part of this dissertation resolves a discrepancy between findings in animal studies and non-invasive human studies. The human auditory cortex appears to have a specific topography of self-speech suppression that is comprised of a mosaic of suppressed and non-suppressed neuronal responses. During vocalization this mosaic of neuronal populations exhibits a stable spatial pattern on the surface of the cortex, which is seen as an averaged suppressed response when recorded from scalp electrodes. In contrast to previous models, this result implies that the auditory cortex is comprised of an intricate network of functionally distinct regions that are not homogenously suppressed by a remote cortical source.

The next chapter of the dissertation delves into a deeper investigation of this network by examining a finer spatial resolution as well as more complex speech stimuli. The results revealed a rich mosaic of language activity, which exhibited distinct and functionally inverse responses at four mm separation. The data provided first evidence for temporal lobe specificity to words as well as self produced speech. In contrast to most reports in the literature, cortical processing in the temporal lobe is not spatially homogenous but is comprised of spatially independent and functionally distinct sub-regions. Nearby sub-regions of cortex can exhibit extremely different functional selectivity to stimuli and even disregard external stimuli. These surprising results contradict classical theories of language processing in the human temporal lobe.

The last chapter of the dissertation focuses on the role of the frontal cortex in the production of speech. Cortical processing was tracked spatially and temporally across cortex, providing a new role for Broca's area in speech production. Broca's area is classically considered to support speech output. The results reveal a temporal propagation from temporal cortex to Broca's area and culminating in motor cortex. Broca's area was not active during actual articulation but was robustly active prior to articulation. The data suggests that Broca's area serves as an interface between perception and production, supporting articulatory preparation but not involved in the actual act of speaking.

The classical theory of language, which was established over a century ago, posits that speech perception is supported by Wernicke's area in the temporal lobe and speech

production is supported by Broca's area in the frontal lobe. This dichotomous model has been readdressed over the past century but many critical questions remained open due to limitations in either temporal or spatial information. This dissertation circumvented these limitations and addressed several issues cardinal to our understanding of the neurobiological basis of language. The results defy century old dogmas and suggest that language is supported by a complex network of independent sub-regions, with Broca's area acting as a mediator between perception and production rather than as the seat of articulation. While much work remains to be done I hope the findings reported here will help advance the field forward.

References

Allen EA, Pasley BN, Duong T, Freeman RD (2007) Transcranial magnetic stimulation elicits coupled neural and hemodynamic consequences. Science 317:1918-1921.

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

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

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

Berger H (1929). Über das Elektrenkephalogramm des Menchen. Archives für Psychiatrie, 87, 527-570.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. J Neurosci 17:353-362.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512-528.

Blumstein S, Baker E, Goodglass H (1977a) Phonological factors in auditory comprehension in aphasia. Neuropsychologia 15(1),19-30.

Blumstein SE, Cooper WE, Zurif EG, Caramazza A (1977b) The perception and production of voice onset time in aphasia. Neuropsychologia 15, 371–383.

Boatman D, Lesser RP, Gordon B (1995) Auditory speech processing in the left temporal lobe: an electrical interference study. Brain and language 51:269-290.

Broca, P. (1861). Remarques sur le siege de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). Bulletins et mémoires de la Société Anatomique de Paris 36, 330–357.

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

Bruns A (2004) Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? Journal of Neuroscience Methods 137:321-332.

Burton MW, Small S, Blumstein SE (2000) The role of segmentation in phonological processing: an fMRI investigation. Journal of Cognitive Neuroscience 12, 679–690.

Canolty R, Soltani M, Dalal SS, Edwards E, Dronkers NF, Nagarajan SS, Kirsch HE, Barbaro NM, Knight RT (2007) Spatiotemporal dynamics of word processing in the human brain. Front Neurosci 1:185-196.

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

Ceponiene R, Cheour M, Näätänen R (1998) Interstimulus interval and auditory event-related potentials in children: evidence for multiple generators. Electroencephalogr Clin Neurophysiol 108:345-354.

Christoffels IK, Formisano E, Schiller NO (2007) Neural correlates of verbal feedback processing: an fMRI study employing overt speech. Human Brain Mapping 28:868-879.
Creutzfeldt O, Ojemann, Lettich (1989) Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. Exp Brain Res 77:476-489.
Crone NE, Miglioretti DL, Gordon B, Lesser RP (1998) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain 121 (Pt 12):2301-2315.
Crone NE, Boatman D, Gordon B, Hao L (2001a) Induced electrocorticographic gamma activity during auditory perception. Brazier Award-winning article, 2001. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology 112:565-582.
Crone NE, Hao L, Hart J, Boatman D, Lesser RP, Irizarry R, Gordon B (2001b) Electrocorticographic gamma activity during word production in spoken and sign language. Neurology 57:2045-2053.
Curio G, Neuloh G, Numminen J, Jousmäki V, Hari R (2000) Speaking modifies voice-evoked activity in the human auditory cortex. Human Brain Mapping 9:183-191.
Damasio AR (1992) Aphasia. New Engl. J. Med. 326, 531-539.
Davis, PA (1939) Effects of acoustic stimuli on the waking human brain. The Journal of Neurophysiology, 2, 494:499.
Démonet JF, Chollet F, Ramsay S, Cadebat D, Nespoulous JL, Wise RJS, Rascol A, Frackowiak RSJ (1992) The anatomy of phonological and semantic processing in normal subjects. Brain 115 (Pt 6):1753-1768.
Dronkers NFA (1996) A new brain region for coordinating speech articulation. Nature, 384, 159–61.
Dronkers NF, Wilkins DP, Van VR, Redfern BB, Jaeger JJ (2004) Lesion analysis of the brain areas involved in language comprehension. Cognition 92, 145–177.
Edwards E, Soltani M, Deouell LY, Berger MS, Knight RT (2005) High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. Journal of Neurophysiology 94:4269-4280.
Edwards E, Srikantan SS, Sarang SD, Canoty RT, Kirsch HE, Barbaro NM, Knight RT (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage 50(1), 291-301.
Eliades S (2005) Dynamics of Auditory-Vocal Interaction in Monkey Auditory Cortex. Cerebral Cortex 15:1510-1523.
Eliades S, Wang X (2008) Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453:1102-1106.
Eliades SJ, Wang X (2003) Sensory-Motor Interaction in the Primate Auditory Cortex During Self-Initiated Vocalizations. Journal of Neurophysiology 89:2194.
Ford JM, Mathalon DH, Heinks T, Kalba S, Faustman WO, Roth WT (2001) Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. The American journal of psychiatry 158:2069-2071.
Flinker A, Chag EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010) Single-trial speech suppression of auditory cortex activity in humans. Journal of Neuroscience. 30, 16643–16650.

Flinker A, Chang EF, Barbaro NM, Berger MS, Knight RT (2011) Sub-centimeter language organization in the human temporal lobe. Brain and Language 117(3), 103-9.
Freeman W (2004) Origin, structure, and role of background EEG activity. Part 1. Analytic amplitude. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology 115:2077-2088.
Fried I, Ojemann GA, Fetz EE (1981) Language-related potentials specific to human language cortex. Science, 212, 353-356.
Friederici AD, Pfeifer E, Hahne A (1993) Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. Brain Res Cogn Brain Res 1(3):183-92.
Friederici AD (2002) Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences, 6, 78-84.
Fu MJ, Daly JJ, Cavuşoğlu MC (2006) A detection scheme for frontalis and temporalis muscle EMG contamination of EEG data. Conf Proc IEEE Eng Med Biol Soc 1:4514-4518.
Geschwind, N (1970) The organization of language and the brain. Science 170, 940-944.
Goncharova II, McFarland DJ, Vaughan TM, Wolpaw JR (2003) EMG contamination of EEG: spectral and topographical characteristics. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology 114:1580-1593.
Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. The Journal of Comparative Neurology 394:475-495.
Hagoort P (2005) On Broca, brain, and binding: a new framework. Trends in Cognitive Sciences 9, 416–423.
Hahne, A and Friederici, AD (1999) Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. J. Cogn. Neurosci.11,194-205.
Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM (2005) Fine-tuning of auditory cortex during speech production. Psychophysiology 42:180-190.
Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393-402.
Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: an MEG study. Journal of cognitive neuroscience 14:1125-1138.
Indefrey P and Levelt WJ (2004) The spatial and temporal signatures of word production components. Cognition, 92, 101–144.
Kaan E, Harris A, Gibson E, Holcomb P (2000) The P600 as an index of syntactic integration difficulty. Language and Cognitive Processes 15(2):159-201.
Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic anomaly. Science 207, 203-205.
Kutas M, Federmeier KD (2000) Electrophysiology reveals semantic memory usein language comprehension. Trends Cogn. Sci. 4, 463-470.
Lane H, Tranel B (1971) The Lombard Sign and the Role of Hearing in Speech. J Speech Hear Res 14:677-709.

Levelt WJM (1983) Monitoring and self-repair in speech. Cognition 14:41-104.
Levelt WJ, Roelofs A, Meyer AS (1999) A theory of lexical access in speech production. Behavioral and Brain Sciences 22(1), 1-75.
Liu J, Newsome WT (2006) Local field potential in cortical area MT: stimulus tuning and behavioral correlations. Journal of Neuroscience 26:7779-7790.
Llorens A, Trébuchon A, Liégeois-Chauvel C, Alario FX (2011) Intra-cranial recordings of brain activity during language production. Frontiers in Psychology 2, 375.
Miller KJ, Leuthardt EC, Schalk G, Rao RP, Anderson NR, Moran DW, Miller JW, Ojemann JG (2007) Spectral changes in cortical surface potentials during motor movement. Journal of Neuroscience 27:2424-2432.
Mohr JP, Pessin MS, Finkelstein S, Funkenstein HH, Duncan GW, Davis KS (1978) Broca aphasia - pathologic and clinical. Neurology, 28, 311-324.
Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R (2005) Coupling between neuronal firing, field potentials, and FMRI in human auditory cortex. Science 309:951-954.
Müller-Preuss P, Ploog D (1981) Inhibition of auditory cortical neurons during phonation. Brain Research 215:61-76.
Mummery CJ, Ashburner J, Scott SK, Wise RJ (1999) Functional neuroimaging of speech perception in six normal and two aphasic subjects. The Journal of the Acoustical Society of America 106:449-457.
Näätänen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24(4):375-425.
Neville H, Nicol JL, Brass A, Forster KI, Garrett MF (1991) Syntactically Based Sentence Processing Classes: Evidence from Event-Related Brain Potentials. Journal of cognitive neuroscience 3(2):151-165.
Numminen J, Salmelin R, Hari R (1999) Subject's own speech reduces reactivity of the human auditory cortex. Neuroscience Letters 265:119-122.
Nunez PL and Srinivasan R (2005) Electric Fields of the Brain: The Neurophysics of EEG., 2nd Edition: New York: Oxford University Press.
Ojemann G, Ojemann J, Lettich E, Berger M (1989) Cortical language localization in left, dominant hemisphere. Journal of Neurosurgery 71:316-326.
Osterhout L, Holcomb PJ (1992) Event-related brain potentials elicited by syntactic anomaly. Journal of Memory and Language 31(6) 785-806.
Papademetris X, Jackowski M, Rajeevan N, Constable RT, Staib LH BioImage Suite: An integrated medical image analysis suite, Section of Bioimaging Sciences, Dept. of Diagnostic Radiology, Yale School of Medicine. http://www.bioimagesuite.org.
Pei X, Leuthardt EC, Gaona CM, Brunner P, Wolpaw JR, Schalk G (2011) Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. Neuroimage 54, 2960–2972.
Petersen S, Fox P, Posner M, Mintun M (1989) Positron Emission Tomographic Studies of the Processing of Singe Words. Journal of Cognitive Neuroscience 1(2):153-170.
Price CJ, Wise RJ, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RS, Friston KJ (1996) Hearing and saying - The functional neuro-anatomy of auditory word processing. Brain 119 (Pt 3):919-931.

Pritchard W (1992) The Brain in Fractal Time: 1/F-Like Power Spectrum Scaling of the Human Electroencephalogram. Int J Neurosci 66:119-129.
Rauschecker JP (1998) Parallel processing in the auditory cortex of primates. Audiol Neurootol 3:86-103.
Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci USA 97:11800-11806.
Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268:111-114.
Ray S, Crone NE, Niebur E, Franaszczuk PJ, Hsiao SS (2008) Neural correlates of high-gamma oscillations (60-200 Hz) in macaque local field potentials and their potential implications in electrocorticography. Journal of Neuroscience 28:11526-11536.
Röder B, Rösler F, Neville HJ (1999) Effects of interstimulus interval on auditory event-related potentials in congenitally blind and normally sighted humans. Neuroscience Letters 264:53-56.
Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nature Neuroscience 2:1131-1136.
Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E (2009) Sequential Processing of Lexical, Grammatical, and Phonological Information Within Broca's Area. Science 326, 445.
Sanai N, Mirzadeh Z, Berger MS (2008) Functional outcome after language mapping for glioma resection. The new england journal of medicine 358:18-27.
Scott SK, Blank CC, Rosen S, Wise RJ (2000) Identification of a pathway for intelligible speech in the left temporal lobe. Brain 12:2400-2406.
Scott SK and Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. Trends in Neurosciences vol. 26 (2) pp. 100-7.
Sohal VS, Zhang F, Yizhar O, Deisseroth K (2009) Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. Nature 459:698.
Sperry R (1950) Neural basis of the spontaneous optokinetic response produced by visual inversion. Journal of comparative and physiological psychology 43:482-489.
Tanji K, Suzuki K, Delorme A, Shamoto H, Nakasato N (2005) High-frequency gamma-band activity in the basal temporal cortex during picture-naming and lexical-decision tasks. Journal of Neuroscience 25:3287-3293.
Towle VL, Yoon H, Castelle M, Edgar JC, Biassou NM, Frim DM, Spire J, Kohrman MH (2008) ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain 131:2013.
Trautner P, Rosburg T, Dietl T, Fell J, Korzyukov OA, Kurthen M, Schaller C, Elger CE, Boutros NN (2006) Sensory gating of auditory evoked and induced gamma band activity in intracranial recordings. NeuroImage 32:790-798.
v. Békésy G (1949) The Structure of the Middle Ear and the Hearing of One's Own Voice by Bone Conduction. The Journal of the Acoustical Society of America 21:217.
Vaden KI, Hickok GS, Halpin HR (2009) Irvine Phonotactic Online Dictionary, Version 1.4*. [Data file] Available from http://www.iphod.com.
Von Holst E, Mittelstaedt M (1950) Das Reafferenzprinzip. Naturwissenschaften 37:464-476.

Wernicke C (1874) The symptom complex of aphasia. A psychological study of an anatomical basis. In Boston Studies in the Philosophy of Science: Proceedings of the Boston Colloquium for the Philosophy of Science, 1966/1968 Volume IV.
Whitham EM, Lewis T, Pope KJ, Fitzgibbon SP, Clark CR, Loveless S, DeLosAngeles D, Wallace AK, Broberg M, Willoughby JO (2008) Thinking activates EMG in scalp electrical recordings. Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology 119:1166-1175.
Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA (2001) Separate neural subsystems within 'Wernicke's area'. Brain 124:83-95.
Wise RJS, Greene J, Buchel C, Scott SK (1999) Brain regions involved in articulation. In: Lancet, pp 1057-1061.
Wise RJS, Chollet F, Hadar U, Friston K, Hoffner E, Frackowiak RSJ (1991) Distribution of cortical neural networks involved in word comprehension and word retrieval. Brain 114 (Pt 4):1803-1817.
Yates A (1963) Delayed auditory feedback. Psychol Bull 60:213-232.
Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY (2008) Transient induced gamma-band response in EEG as a manifestation of miniature saccades. Neuron 58:429-441.
Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846-849.
Zatorre R, Meyer E, Gjedde A, Evans A (1996). PET Studies of Phonetic Processing of Speech: Review, Replication, and Reanalysis. Cerebral Cortex 6(1):21-30.