

Published in final edited form as:

Neuroimage. 2014 May 1; 91: 91–98. doi:10.1016/j.neuroimage.2014.01.003.

Action planning and predictive coding when speaking

Jun Wang¹, Daniel H. Mathalon^{2,3}, Brian J. Roach², James Reilly⁴, Sarah Keedy⁵, John A. Sweeney¹, and Judith M. Ford^{2,3,*}

¹Department of Psychiatry, University of Texas Southwestern, Dallas, TX, 75390

²San Francisco VA Medical Center, San Francisco, CA, 94121

³Department of Psychiatry, University of California, San Francisco, CA, 94121

⁴Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, 60611

⁵Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, 60637

Abstract

Across the animal kingdom, sensations resulting from an animal's own actions are processed differently from sensations resulting from external sources, with self-generated sensations being suppressed. A forward model has been proposed to explain this process across sensorimotor domains. During vocalization, reduced processing of one's own speech is believed to result from a comparison of speech sounds to corollary discharges of intended speech production generated from efference copies of commands to speak. Until now, anatomical and functional evidence validating this model in humans has been indirect. Using EEG with anatomical MRI to facilitate source localization, we demonstrate that inferior frontal gyrus activity during the 300ms before speaking was associated with suppressed processing of speech sounds in auditory cortex around 100ms after speech onset (N1). These findings indicate that an efference copy from speech areas in prefrontal cortex is transmitted to auditory cortex, where it is used to suppress processing of anticipated speech sounds. About 100ms after N1, a subsequent auditory cortical component (P2) was not suppressed during talking. The combined N1 and P2 effects suggest that although sensory processing is suppressed as reflected in N1, perceptual gaps are filled as reflected in the lack of P2 suppression, explaining the discrepancy between sensory suppression and preserved sensory experiences. These findings, coupled with the coherence between relevant brain regions before and during speech, provide new mechanistic understanding of the complex interactions between action planning and sensory processing that provide for differentiated tagging and monitoring of one's own speech, processes disrupted in neuropsychiatric disorders.

Keywords

Corollary discharge; efference copy; IFG; STG

*Corresponding author: Judith M. Ford, Ph.D. Psychiatry Service (116D) San Francisco VA Medical Center 4150 Clement Street San Francisco, CA 94121 Judith.ford@ucsf.edu 415 221-4810, extension 4187 415 750-6622 (fax).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest statement
There are no conflicts of interest.

Introduction

Perception would be relatively uncomplicated if an animal never moved, but moving introduces a cascade of challenges for sensory processing (Crapse and Sommer, 2008). For example, the nematode must decide if the pressure on its nose comes from its own forward swimming action or from an approaching predator (Crapse and Sommer, 2008). For humans, fully processing our own speech while talking would divert processing resources from more important and unexpected environmental events. Nature has devised a forward model system to deal with these challenges that minimizes processing of sensations that result from our own actions while tagging them as coming from “self”.

In addition to primary motor commands to move, the forward model proposes two secondary components: An efference *copy* of the plan or command to move, and a corollary discharge of the expected sensation resulting from the movement (Miall and Wolpert, 1996) (Figure 1). In the case of speaking, the corollary discharge of expected speech sounds is effectively compared to the actual sound (Miall and Wolpert, 1996). One benefit to this process is that when there is a mismatch between the heard and intended speech sounds, future speech production can be corrected to more closely approach the desired form.

The neurophysiology of the corollary discharge during vocalization has been studied across the animal kingdom, from songbirds (Keller and Hahnloser, 2009), crickets (Poulet and Hedwig, 2006) and bats (Suga and Shimozawa, 1974) to non-human primates (Eliades and Wang, 2003; Eliades and Wang, 2005; Eliades and Wang, 2008) and humans (Chen et al., 2011; Creutzfeldt et al., 1989). In humans, suppression during talking is seen as a marked reduction in N1 of the EEG-based event-related potential (ERP), and M100 of the MEG-based response (Curio et al., 2000; Ford et al., 2001; Heinks-Maldonado et al., 2005; Houde et al., 2002; Tian and Poeppel, 2013) both emanating from auditory cortex (Hari et al., 1987; Krumbholz et al., 2003; Ozaki et al., 2003; Pantev et al., 1996; Reite et al., 1994; Sams et al., 1985). This has been shown with intracranial recordings from the cortical surface of patients receiving surgery for epilepsy (Chen et al., 2011; Flinker et al., 2010). These findings are believed to reflect an active suppression of auditory cortical responses during talking via a forward model system (Curio et al., 2000; Ford et al., 2001; Heinks-Maldonado et al., 2005; Houde et al., 2002). Less is known about the neurophysiology of action planning preceding vocalization and its relation to subsequent sensory suppression, although in song birds and humans, pre-song (Keller and Hahnloser, 2009) and pre-speech (Ford et al., 2002; Ford et al., 2007b) activity has been suggested to reflect the action of the efference copy. Even less is known about our ability to simultaneously suppress and preserve perception of our speech while talking. Essentially, the question is: how do we still seem to “hear” our own speech, as we obviously do, when the sensory processing of our own speech is markedly suppressed? To address this, we assessed speech-dependent neural activity in an ERP component following N1, namely P2.

P2 is positive component typically seen in conjunction with N1, but independent from it. The literature on the auditory P2 is sparse, perhaps because it has been difficult to relate it to specific processes (for review see (Crowley and Colrain, 2004)). The literature is mixed regarding whether P2 is suppressed to self-generated sounds, with some studies showing suppression (Greenlee et al., 2013; Houde et al., 2002; Knolke et al., 2013) and some not (Baess et al., 2008; Behroozmand and Larson, 2011; Martikainen et al., 2005) and many not reporting P2 data at all (e.g., (Ford et al., 2007a)).

To fully examine the real-time action of forward model systems during human vocalization, techniques are needed with both high temporal and spatial resolution, to examine how speech production and speech perception brain regions interact and influence each

other(Price et al., 2011) before and during sensory processing. Previous support for the forward model in humans has been indirect and anatomically imprecise(Ford et al., 2002; Ford et al., 2007b), has suffered from low temporal resolution (Price et al., 2011; Zheng et al., 2010; Zheng et al., 2013), and has come from invasive studies of small groups of patients with epilepsy with non-ideal electrode placement and abnormal brain neurophysiology (Chang et al., 2013; Chen et al., 2011). The MRI facilitated source localization of EEG recordings allows elements of the forward model to be examined and established with considerably more precision. Indeed, several studies have utilized individual MRI to facilitate or guide source localization (Curio et al., 2000; Houde et al., 2002; Ventura et al., 2009). In the source coherence analysis presented here, we also used MRI-guided source localization. We examined the neural source of the efference copy before speech onset, and its relation to activity in auditory cortex before and after speech onset. We examined subsequent neural activity that may underlie seamless perception of speech during talking. To address the neural underpinnings of these different aspects of speech, we compared conditions in which subjects talked to conditions in which they listened to a recording of their speech production.

Materials and Methods

Subjects

Thirty-six healthy subjects (15 female, aged 16-56) participated in the experiment. All participants were right handed and had normal or corrected-to-normal vision. Participants provided informed consent prior to testing and were paid \$15/h for their participation. The Institutional Review Board of the University of Illinois at Chicago approved the study.

Procedure

Participants completed the Talking/Listening paradigm, as described previously (Ford et al., 2010) (Figure 1) using Presentation software (www.neurobs.com/presentation). For the Talking task, participants pronounced short (<300ms), sharp vocalizations of the phoneme “ah” repeatedly in a self-paced manner, about every 1-2s, for 180s. Participants vocalized about every 2s (median ah-onset asynchrony: 2000.8 ms, inter-quartile range: 1631.3-2269ms). The speech was recorded using a microphone and transmitted back to subjects through headphones in real time (zero delay). In the Listening task, the recording from the Talking task was played back, and participants were instructed simply to listen. Sound intensity was kept the same in Talking and Listening tasks for each participant by ensuring that a 1000Hz tone (generated by a Quest QC calibrator) produced equivalent dB intensities when delivered through earphones during the tone's generation (Talking task) and during its playback (Listening task). Vocalizations were processed offline to identify vocalization onset(Ford et al., 2010). Trigger codes were inserted into the continuous EEG file at these onsets to time-lock speech epochs and EEG data over trials.

EEG recording

EEG was collected from 64 sintered Ag/AgCl sensors (Quik-Cap, Compumedics Neuroscan, Charlotte, NC) with a forehead ground and nose reference, and impedance was kept below 5k Ω . Electrodes placed at the outer canthi of both eyes, and above and below the right eye, were used to record vertical and horizontal electro-oculogram data. EEG data were digitized and recorded at 1000Hz continuously throughout testing. The 3D locations of all electrodes and three major fiducials (nasion, left, and right peri-auricular points) were digitized using a 3Space Fastrack 3D digitizer (Polhemus, Colchester, VT).

MRI Scans

For EEG-MRI coregistration, structural T1 MRI scans were obtained using a 3T GE Signa Scanner with an 8 channel headcoil and an enhanced fast gradient echo 3-dimensional sequence (166 slices, TR 6.98 ms, TE 2.848 ms, flip angle 8 degrees, voxel size 1.016 mm × 1.016mm × 1.2mm). Head meshes were created from each participant's MRI scan that were coregistered to a 3D map of individual electrode positions via the three fiducial markers.

EEG Analyses

Raw EEG data were filtered using a 1Hz high-pass filter using EEGLAB toolbox (Delorme and Makeig, 2004), then subjected to Fully Automated Statistical Thresholding for EEG artifact Rejection (FASTER) using a freely distributed toolbox (Nolan et al., 2010). After artifact correction, all further analyses were carried out using SPM8 for MEG/EEG (www.fil.ion.ucl.ac.uk/spm/) and FieldTrip (Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands: <http://www.ru.nl/neuroimaging/fieldtrip/>). EEG data were low pass filtered at 50Hz and transformed to an average reference. Data were epoched from -800 to 800 ms with respect to the onset of each “ah” and baseline corrected using data from the -800 to -500ms epoch preceding vocalization. ERP averages were generated using a robust averaging approach (Wager et al., 2005) included in SPM8. Inspection of the grand average ERP waveform indicated three components: N1 peaked at ~100ms after speech onset, P2 peaked at ~200ms after speech onset, and a slow negative component before speech onset from -300 to 0 ms (Figure 2). Mean amplitudes of these components were extracted at the anterior locus from which each of these three components were robust (collapsed across 26 electrodes surrounding FCz, marked with red circle in Figure 2): N1: 80-120ms; P2: 170-210 ms; pre-speech: -300-0 ms.

Source estimates

Source reconstruction analyses were conducted to localize neural generators producing ERP components (Pre-Speech-Onset, N1, P2). The forward model's lead field was computed using the boundary element method EEG head model available in SPM8 constrained by individual's MRI-determined brain anatomy. Effects in the low frequency band (1-15 Hz) were source-localized by computing the inverse model lead field. Source estimates for each subject were computed on average ERP waveforms for Talking and Listening using the Pre-Speech Onset period from -300 to 0ms, the Post-Speech-Onset period from 80 to 120ms centered on the N1 peak and the Post-Speech-Onset period from 170 to 210ms centered on the P2 peak. An epoch -800 to -500ms was also evaluated to control for potential task differences in Baseline activity. The source estimate computation was implemented on the canonical mesh using multiple sparse priors (Friston et al., 2008) under group constraints (Litvak and Friston, 2008). A two-way repeated measures ANOVA design was used in SPM for Task (Talking, Listening) and Period (Baseline, pre-speech, N1, P2). The ANOVA was carried out on voxels over the whole brain; significant clusters of voxels with task differences were identified using a FWE-corrected cluster threshold of $p < 0.05$, after voxel level thresholding at $p < 0.001$ in order to correct for multiple comparisons.

Pre-speech-Onset Source Coherence between IFG and Auditory Cortex

To determine whether the Pre-Speech activity in IFG was synchronized with activity in auditory cortex before speech onset, we calculated phase-locking value (PLV or phase coherence) of the EEG between these two regions. First, we extracted source activity from bilateral IFG and bilateral STG, with each region defined as a sphere with a 12mm radius ROI centered on the significant clusters identified in the Task X Period interaction included in aforementioned ANOVA model: IFG ($x=-48, y=26, z=-2$, and $x=52, y=24, z=0$) and STG ($x=-62, y=-32, z=10$; $x=58, y=-34, z=8$). Next, we imported the single trial source activity in

each ROI to FieldTrip for subsequent analysis where the single trial source data were transformed from the time to the frequency domain using a fast Fourier transform (FFT), with a frequency resolution of $\sim 3.33\text{Hz}$. At each frequency step, the normalized phase difference between speech production and auditory cortical sources was calculated and averaged across trials to derive the PLV for that frequency (Lachaux et al., 1999).

For each subject, we also calculated a phase locking statistic (PLS) (Lachaux et al., 1999) to test the significance of PLVs. This was performed using a non-parametric permutation approach, which shuffled the trial order of the auditory cortex source trials while keeping the same trial order for the IFG area. PLV was computed for each permutation and repeated 200 times providing a distribution of values from which the PLS was derived. The original PLV at each frequency for each subject was considered statistically significant if it was beyond the 95% percentile of the PLS distribution.

Greenhouse-Geisser corrections for non-sphericity were used when appropriate.

Results

ERP analyses

The ERP responses to speech sound onset are shown in Figure 2. On the right, we show the grand average ERPs recorded from the frontal-central midline scalp site (FCz) where signals generated in auditory cortex are readily detected. On the left, we show the scalp maps representing activity at all recording sites. A repeated measures ANOVA was used to assess the effects of Task (Talking vs. Listening) on the amplitude of the ERP Components (Pre-speech from -300 to 0ms, N1, P2). There was a significant Task by ERP Component interaction ($F(2,70)=17.03$, $p<0.001$), a main effect of ERP Component ($F(2,70)=39.55$, $p<0.001$), and a main effect of Task ($F(1,35)=4.34$, $p=0.045$). The interaction was followed-up using paired t-tests to compare the effects of Task for each ERP Component. Pre-speech negativity was larger during Talking than Listening ($t(35)=2.55$, $p=0.015$), N1 was larger during Listening than Talking ($t(35)=6.26$, $p<0.001$), and P2 did not differ between Talking and Listening tasks ($t(35)=1.40$, $p=0.169$).

Source estimates

Source estimates derived from the ERPs revealed source differences between Talking and Listening mainly in low frequencies (1-15Hz). Indeed, the source estimates using EEG filtered between 15-50Hz did not differ between Talking and Listening. Therefore, source analyses were computed focusing on EEG activity between 1-15Hz, similar to our previous methods (Ford et al., 2007a) and those of others (Knollea et al., 2013).

There was a Task x Period interaction ($F's(3,280)>12.10$; FWE corrected to $p<.05$ at voxel level), but no main effect of Task or Period. There was no effect of Task during the Baseline or P2 periods. However, there were significant interactions of Task x Period (Pre-Speech-Onset vs. N1) in bilateral inferior frontal gyrus (p values <0.018 , FWE corrected), bilateral sensorimotor mouth area (p values <0.038 , FWE corrected) and bilateral superior temporal gyrus (p values <0.042 , FWE corrected) (Figure 3a). These interactions were followed up with pairwise comparisons, using the significant interaction cluster image from Task x Period (Pre-Speech-Onset vs. N1) as an inclusive mask. These tests showed that during the Pre-Speech-Onset period, there was greater activity during Talking than Listening in bilateral IFG (Red color in Figure 3a, p values <0.018 , FWE corrected) and in bilateral sensorimotor mouth area (Red color in Figure 3a, p values <0.038 , FWE corrected). During the N1 period, there was less activity during Talking than Listening in bilateral STG (Green color in Figure 3a, p values <0.048 , FWE corrected).

Because P2 was not affected by Task, we collapsed across Talking and Listening to assess its source. We found stronger activity in bilateral STG during the P2 compared to the baseline epoch (Blue color, Figure 3b; p values <0.040 , FWE corrected). As can be seen in Figure 3b, N1 and P2 sources were highly overlapping.

Pre-speech-Onset Source Coherence between IFG and Auditory Cortex

Using PLV and PLS, we determined that during both Talking and Listening, coherence values before speech onset exceeded the $p<0.05$ threshold (Figure 4). In addition, we used a 3-way repeated measures ANOVA for the factors of Task (Talking vs. Listening), Frequency (3.33, 6.66, 9.99, 13.32 Hz), and Hemisphere (left IFG area to left auditory cortex vs. right IFG area to right auditory cortex). There was a significant Task \times Frequency interaction ($F(1.45,50.85)=8.14$, $p=0.002$, $\epsilon=0.484$), but no interactions involving Hemisphere, suggesting effects were similar in both hemispheres. Post-hoc pairwise comparisons showed Task effects in the lower but not higher frequencies due to greater pre-speech source phase coherence between auditory cortex and IFG during Talking than Listening in the delta/theta frequency range (3.33Hz: $t(35)=3.77$, $p=0.001$; 6.66Hz: $t(35)=3.09$, $p=0.004$) but not in higher frequencies (9.99Hz: $t(35)=1.23$, $p=0.225$; 13.32Hz: $t(35)=1.89$, $p=0.067$) (Figure 4).

Relationship between pre-speech activity and N1 suppression

Pre-speech task differences in both source activity and source phase coherence (PLV) were correlated with task differences in N1 (N1 suppression). The task difference in pre-speech source activity was extracted from IFG and the sensorimotor mouth area where significant source estimates were identified in SPM-converted images for each subject using MarsBar toolbox for SPM8 (<http://marsbar.sourceforge.net/>) (Brett et al., 2002)) and normalized through log transformation of the $\text{Source}_{\text{talk}}/\text{Source}_{\text{listen}}$ ratio. The task difference in pre-speech source phase coherence in the delta/theta band was calculated by subtracting Listening from Talking task PLV, producing positive values in cases of greater phase coherence during the Talking task. The task difference in the N1 ERP was quantified by subtracting N1 amplitude during the Listening task from N1 amplitude during the Talking task, yielding positive values in cases of N1 suppression.

N1 suppression was correlated with the degree of increased Pre-Speech Onset activity in left ($r(36)=0.47$, $p=0.004$) and right IFG ($r(36)=0.48$, $p=0.003$) during Talking (Figure 5a). There were also significant correlations between activity in bilateral IFG during Pre-Speech-Onset and activity in STG during Post-Speech-Onset source activity, but only during the Talking task (Left IFG to left A1: $r(36)=-0.33$, $p=0.049$; Right IFG to right A1: $r(36)=-0.39$, $p=0.018$). In addition, differential pre-speech source phase coherence (Talking-Listening) between the IFG areas and auditory cortex (averaged in delta/theta band and across two inter-area pairings) was correlated with N1 suppression ($r(36)=0.36$, $p=0.03$) (Figure 5b). That is, N1 suppression to the speech sound was greater in subjects who had greater pre-speech activity in left hemisphere IFG areas (and their right hemisphere homologue) and stronger phase coherence before speech onset between auditory cortex and the IFG areas. There was no significant correlation between Pre-Speech Onset activity in the sensorimotor mouth area and N1 suppression.

Relationship between N1 and P2

To understand the interdependence of P2 on N1, which would be high if P2 simply reflected a later serial stage of auditory processing, we computed bivariate correlations between N1 and P2 ERP components during Talking and Listening, separately. Larger N1 amplitudes were associated with larger P2 amplitudes during Listening ($r(36)=-0.40$, $p=0.015$), but not during Talking ($r(36)=0.08$, $p=0.656$). A repeated measures ANCOVA was used to test the

slopes of these two associations and showed significant differences ($F_{1,33}=4.40$, $p=0.044$), suggesting N1 and P2 were more related to each other during Listening than during Talking and that they reflect different processes during Talking. This is consistent with the observation that N1 but not P2 is suppressed during Talking. To further tease apart N1 and P2, we calculated the bivariate correlations between pre-speech source activity in IFG and P2 during Talking. Correlations between IFG source activity and P2 were not significant (Left IFG to P2: $r(36)=0.16$, $p=0.348$; Right IFG to P2: $r(36)=0.24$, $p=0.167$), while these relationships were significant for N1, as detailed above.

Discussion

Our results provide direct anatomical and functional neurobiological support for the forward model of motor and sensory integration during speech in humans. Before discussing the details and implications of our findings, we summarize the data as they map onto the model in Figure 1, starting with the motor plan to speak followed by the brain's response to the resulting sounds.

First, we observed a slow negative-going potential preceding speech onset. We localized this to the posterior ventrolateral frontal lobe bilaterally, illustrated as the “thought bubble” in Figure 1. We observed two foci of activity in this area, encompassing two key regions involved in speech production: The inferior frontal gyrus (IFG), including Broca's area, and the more posterior sensorimotor mouth area, involved in articulation. Only activity in IFG was correlated with suppression of activity in auditory cortex, suggesting it is the neural instantiation of the efference copy of the forward model, illustrated by the bent arrow in Figure 1. According to this figure, the efference copy deposits a corollary discharge of the intended sound in auditory cortex (blue burst) against which the actual sound (green burst) is compared. Importantly, pre-speech activity in sensorimotor mouth area of cortex was not related to N1 suppression. Instead, we suggest that rather than being the “efference copy” of the command to speak, it represents the instantiation of the motor command itself.

Second, we observed coherent neural activity between IFG and auditory cortex before speech onset (perhaps carried on the bent arrow in Figure 1). This activity was stronger during talking than listening, and it too was related to subsequent N1 suppression. The pre-speech coherence between these regions may represent the transfer and instantiation of the efference copy before speech onset (Ford and Mathalon, 2004), and our source-localized EEG data demonstrate this important component of the forward model using neurophysiological data. Moreover, pre-speech source waveforms latencies analysis indicated only during talking, activity in auditory cortex is following the activity in IFG and sensorimotor area in time.

Third, we observed ~80% suppression of neural activity in auditory cortex in response to the speech sound itself, during speaking relative to listening. N1 was localized to the bilateral superior temporal gyrus (STG) of auditory cortex. This remarkable reduction of N1 during talking compared to listening is consistent with single unit data from human (Creutzfeldt et al., 1989) and non-human primates (Eliades and Wang, 2003; Eliades and Wang, 2005; Eliades and Wang, 2008), and confirms that responses in auditory cortex to our own speech are diminished but not completely silenced. Following N1, there was a P2 auditory component that was also localized to auditory cortex. However, it was not suppressed nor was it related to N1 during talking or to pre-speech IFG activity. We speculate that the preserved P2 component reflects a perceptual filling of gaps created by sensory suppression to preserve perceptual experience during speaking.

Thus, our findings suggest that before speech sounds are uttered, the brain generates motor-related commands. The primary motor command is sent to the articulators in sensorimotor cortex, enabling speech. A second set of commands is sent to auditory cortex, the efference copy heralding the impending arrival of specific speech sounds that provides the basis for the development of a corollary discharge to suppress, but not silence, sensory processing of expected sounds and possibly for tagging them as coming from “self”. When the pitch of the sound does not match the expected sound, the auditory response is less reduced (Eliades and Wang, 2008; Heinks-Maldonado et al., 2005; Keller and Hahnloser, 2009), and the sound may seem less self-like (Heinks-Maldonado et al., 2007). In lower animals, the forward model may serve both to avoid end-organ fatigue, as in the cricket stridulating (Poulet and Hedwig, 2006), and to identify the source of the sensation, as in the bat emitting sonar signals in a cave (Suga and Shimozawa, 1974). Although the data presented here do not allow us to distinguish between the suppression and tagging functions of the forward model, in humans, the forward model may serve both to tag sensations as coming from self and reduce auditory cortical activity, potentially enabling better monitoring of external sounds while talking.

N1 and P2 were both localized to bilateral STG as reported previously (Bosnyak et al., 2004; Shahin et al., 2007), although P2 sources may be more medial and anterior to N1 (Ross and Tremblay, 2009). More important are the different roles N1 and P2 reflect during talking, with potential parallels to saccadic suppression in the visual system. During saccades, input is blurred and sensory sensitivity reduced (Knöll et al., 2011), yet the gap is filled to maintain continuous perceptual experience (Burr and Morrone, 2012). The reduced processing of visual input depends on the preparation and execution of a saccade; it is not observed when stimuli are moved across the visual field in a manner paralleling effects during saccades. N1 and P2 may be reflecting similar processes, with P2 reflecting the filling of gaps in sensation resulting from suppression during talking, a process that would appear to occur early and in sensory or adjacent primary association cortex. Such a process could serve to keep the flow of perceptual experience continuous. Studies directly manipulating elements of the forward model and assessing perception during talking are needed to further delineate the different processing elements indicated by N1 and P2 during talking, and the mechanism by which the gap is filled following suppression so that we have the experience of hearing what we say.

Although we cannot prove that the pre-speech activity causes sensory suppression, the fact that activity in IFG precedes suppression in auditory cortex and is related to subsequent suppression is consistent with the assumed temporal order of the efference copy and corollary discharge illustrated in Figure 1. The relatively slowly incrementing nature of the negative-going pre-speech activity in IFG reflecting speech preparation process in the 300ms preceding speech onset might reveal the accumulation and coordination of neural computations related to action planning and preparing sensory systems for their expected consequences. The correlation between pre-speech activity in IFG and subsequent suppression of the auditory cortical response to the speech sound, and the functional connection between IFG and auditory cortex revealed in the coherence analysis, provide strong support for the forward model and clarify its neural sources. In particular, based on these data, we propose that a copy of the motor command to speak, or the efference copy, is transferred from the frontal lobe to auditory cortex, where it is used to suppress the auditory response to the spoken sound as represented by the reduced N1 amplitude. This efference copy is independent of the motor command in the sensorimotor mouth area, which was not related to subsequent suppression in auditory cortex. Finally, the use of source coherence to assess frontal-temporal functional connectivity avoided volume conduction confounds (Hoechstetter et al., 2004), which might have affected previous reports of frontal-temporal connectivity (Ford and Mathalon, 2004), thereby allowing us to confirm the

interdependence of neural activity in the specific brain regions of interest. To the best of our knowledge, this is the first study inspecting synchronization dynamics at the source level during human speech. In addition, the significant PLVs during the listen condition may reflect long-range synchrony in the brain, perhaps reflecting preparation, or active listening, for upcoming sound.

Previous reports of responses to both artificial (Suppes et al., 1998) and naturally spoken sentences (Luo and Poeppel, 2007) suggests speech processing and tracking are instantiated mainly via phase locking of low-frequency (<10Hz) neural activity (Luo and Poeppel, 2007). Source phase coherence between IFG and STG and source activity in IFG was also observed in low frequencies prior to speech onset. Similar low frequency activity over bilateral prefrontal cortex has been reported before overt reading (Gehrig et al., 2012) and also when subjects intended to speak (Carota et al., 2010), but this activity had not previously been linked to subsequent changes in sensory processing. Overall, that both speech planning and processing involve neural activity in the theta band suggests that both processes involve similar neural mechanisms (Greenberg and Ainsworth, 2006; Luo and Poeppel, 2007), and that the activity has a temporal resolution of 125-250ms (one cycle of theta band activity) (Drullman et al., 1994; van der Horst et al., 1999). This is roughly the duration of the speech sound “ah” in this paper and the mean syllable length across languages (Greenberg and Ainsworth, 2006). Whether the periodicity of synchrony varies with the duration of speech remains to be addressed.

It is important to mention that the frontal-temporal source coherence reported in the low-frequencies is different from frontal-temporal source coherence in gamma band we reported in Chen et al (Chen et al., 2011). This may due to several factors. First, different measurements were used. Intracranial recordings provide high enough signal-to-noise ratio to allow analysis of synchrony in frequencies with smaller signals, such as gamma. With scalp recordings, high frequency gamma activity is usually affected by miniature saccades (Yuval-Greenberg et al., 2008) and muscle artifact from ongoing speech. Therefore, our study focuses on lower frequencies between 1-15Hz, as we (Ford et al., 2007a) and others (Knolke et al., 2013) have done with scalp recorded data. Second, in Chen et al (Chen et al., 2011), we hypothesized frontal-temporal source coherence in the gamma band and did not evaluate low-frequency results. Third, given the time-frequency method we used in Chen et al in which the wavelet family was defined by a constant ratio ($f_0/\sigma_f=7$), a wider range of window duration would have been needed for evaluating lower frequencies (e.g. 557.4ms for 4Hz). This wide window would be beyond the duration of the pre-speech period and distort our calculation of pre-speech synchrony. In the current paper, we did a fast Fourier transform (FFT) in the whole 300ms pre-speech window, with a frequency resolution of ~3.33Hz, enabling us to assess the source coherence in frequency band as low as 3.33Hz. Processing of both artificial (Suppes et al., 1998) and naturally spoken sentences (Luo and Poeppel, 2007) is instantiated mainly via phase locking of low-frequency (<10Hz) neural activity (Luo and Poeppel, 2007).

In conclusion, while the efference copy and corollary discharge mechanisms have been elegantly studied across the animal kingdom, these data provide anatomic and temporal specificity to both the neural origin of the efference copy and corollary discharge, and their functional integration and effects in humans, using non-invasive methods. Our findings describe the mechanisms responsible for the dramatic ability of the brain to reduce the redundancy of processing predictable sensory consequences of self-initiated behavior in real time. Even though early sensory processing appears robustly suppressed, via potential integration of residual sensory information and the corollary discharge, perceptual experience may be “made whole” to account for the marked discrepancy between a near 80% N1 suppression and the preserved P2 component and perceptual experience of clearly

hearing what one says. This approach can be readily translated to studies of neuropsychiatric populations in whom abnormalities in forward model mechanisms may be associated with misperceptions of origins of thoughts and actions with regard to whether they originate from oneself (Feinberg, 1978; Ford et al., 2007b; Frith, 1987).

Acknowledgments

This work was supported by grants from NIMH MH077862, MH085485, and MH058262. We thank Brett Clementz for comments on an early draft.

REFERENCES

- Baess P, Jacobsen T, Schröger E. Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology*. 2008; 70:137–143. [PubMed: 18627782]
- Behroozmand R, Larson CR. Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*. 2011; 12:1–10. [PubMed: 21208416]
- Bosnyak DJ, Eaton RA, Roberts LE. Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex*. 2004; 14:1088–1099. [PubMed: 15115745]
- Brett, M.; Anton, J-L.; Valabregue, R.; Poline, J-B. Region of interest analysis using an SPM toolbox. In: NeuroImage, V., editor. 8th International Conference on Functional Mapping of the Human Brain; Sendai, Japan. 2002. No 2.
- Burr DC, Morrone MC. Constructing stable spatial maps of the world. *Perception*. 2012; 41:1355–1372. [PubMed: 23513621]
- Carota F, Posada A, Harquel S, Delpuech C, Bertrand O, Sirigu A. Neural dynamics of the intention to speak. *Cerebral Cortex*. 2010; 20:1891–1897. [PubMed: 20008453]
- Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF. Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc Natl Acad Sci U S A*. 2013; 110:2653–2658. [PubMed: 23345447]
- Chen CM, Mathalon DH, Roach BJ, Cavus I, Spencer DD, Ford JM. The Corollary Discharge in Humans Is Related to Synchronous Neural Oscillations. *J Cogn Neurosci*. 2011; 23:2892–2904. [PubMed: 20946054]
- Crapse TB, Sommer MA. Corollary discharge across the animal kingdom. *Nature Reviews. Neuroscience*. 2008; 9:587–600.
- Creutzfeldt O, Ojeman G, Lettich E. Neuronal activity in the human lateral temporal lobe. II Responses to the subject's own voice. *Experimental Brain Research*. 1989; 77:476–489. [PubMed: 2806442]
- Crowley KE, Colrain IM. A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology*. 2004; 115:732–744. [PubMed: 15003751]
- Curio G, Neuloh G, Numminen J, Jousmaki V, Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping*. 2000; 9:183–191. [PubMed: 10770228]
- Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*. 2004; 134:9–21. [PubMed: 15102499]
- Drullman R, Festen JM, Plomp R. Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am*. 1994; 95:1053–1064. [PubMed: 8132899]
- Eliades SJ, Wang X. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *Journal of Neurophysiology*. 2003; 89:2194–2207. [PubMed: 12612021]
- Eliades SJ, Wang X. Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cerebral Cortex*. 2005; 15:1510–1523. [PubMed: 15689521]
- Eliades SJ, Wang X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*. 2008; 453:1102–1106. [PubMed: 18454135]

- Feinberg I. Efference copy and corollary discharge: implications for thinking and its disorders. *Schizophrenia Bulletin*. 1978; 4:636–640. [PubMed: 734369]
- Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT. Single-trial speech suppression of auditory cortex activity in humans. *Journal of Neuroscience*. 2010; 30:16643–16650. [PubMed: 21148003]
- Ford JM, Gray M, Faustman WO, Roach BJ, Mathalon DH. Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology*. 2007a; 44:522–529. [PubMed: 17565658]
- Ford JM, Mathalon DH. Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research*. 2004; 38:37–46. [PubMed: 14690769]
- Ford JM, Mathalon DH, Heinks T, Kalba S, Roth WT. Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry*. 2001; 158:2069–2071. [PubMed: 11729029]
- Ford JM, Mathalon DH, Whitfield S, Faustman WO, Roth WT. Reduced communication between frontal and temporal lobes during talking in schizophrenia. *Biol Psychiatry*. 2002; 51:485–492. [PubMed: 11922884]
- Ford JM, Roach BJ, Faustman WO, Mathalon DH. Synch before you speak: auditory hallucinations in schizophrenia. *Am J Psychiatry*. 2007b; 164:458–466. [PubMed: 17329471]
- Ford JM, Roach BJ, Mathalon DH. Assessing corollary discharge in humans using noninvasive neurophysiological methods. *Nat Protoc*. 2010; 5:1160–1168. [PubMed: 20539291]
- Friston K, Harrison L, Daunizeau J, Kiebel S, Phillips C, Trujillo-Barreto N, Henson R, Flandin G, Mattout J. Multiple sparse priors for the M/EEG inverse problem. *Neuroimage*. 2008; 39:1104–1120. [PubMed: 17997111]
- Frith CD. The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine*. 1987; 17:631–648. [PubMed: 3628624]
- Gehrig J, Wibral M, Arnold C, Kell CA. Setting up the speech production network: how oscillations contribute to lateralized information routing. *Frontiers in psychology*. 2012; 3:169. [PubMed: 22685442]
- Greenberg, S.; Ainsworth, W. *Listening to speech: an auditory perspective*. Erlbaum; New Jersey: 2006.
- Greenlee JDW, Behroozmand R, Larson CR, Jackson AW, Chen F, Hansen DR, Oya H. Sensory-Motor Interactions for Vocal Pitch Monitoring in Non-Primary Human Auditory Cortex. *PLoS One*. 2013; 8:e60783. [PubMed: 23577157]
- Hari R, Pelizzzone M, Makela JP, Hallstrom J, Leinonen L, Lounasmaa OV. Neuromagnetic responses of the human auditory cortex to on- and offsets of noise bursts. *Audiology*. 1987; 26:31–43. [PubMed: 3593099]
- Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM. Fine-tuning of auditory cortex during speech production. *Psychophysiology*. 2005; 42:180–190. [PubMed: 15787855]
- Heinks-Maldonado TH, Mathalon DH, Houde JF, Gray M, Faustman WO, Ford JM. Relationship of Imprecise Corollary Discharge in Schizophrenia to Auditory Hallucinations. *Arch Gen Psychiatry*. 2007; 64:286–296. [PubMed: 17339517]
- Hochstetter K, Bornfleth H, Weckesser D, Ille N, Berg P, Scherg M. BESA source coherence: a new method to study cortical oscillatory coupling. *Brain Topography*. 2004; 16:233–238. [PubMed: 15379219]
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*. 2002; 14:1125–1138. [PubMed: 12495520]
- Keller GB, Hahnloser RH. Neural processing of auditory feedback during vocal practice in a songbird. *Nature*. 2009; 457:187–190. [PubMed: 19005471]
- Knöll J, Binda P, Morrone MC, Bremmer F. Spatiotemporal profile of peri-saccadic contrast sensitivity. *Journal of Vision*. 2011; 11:1–12.
- Knollea F, Schröger E, Kotza SA. Prediction errors in self- and externally-generated deviants. *Biological Psychology*. 2013; 92:410–416. [PubMed: 23246535]

- Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lutkenhoner B. Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*. 2003; 13:765–772. [PubMed: 12816892]
- Lachaux J, Rodriguez E, Martinerie J, Varela F. Measuring phase synchrony in brain signals. *Human Brain Mapping*. 1999; 8:194–208. [PubMed: 10619414]
- Litvak V, Friston K. Electromagnetic source reconstruction for group studies. *Neuroimage*. 2008; 42:1490–1498. [PubMed: 18639641]
- Luo H, Poeppel D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*. 2007; 54:1001–1010. [PubMed: 17582338]
- Martikainen MH, Kaneko K, Hari R. Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex*. 2005; 15:299–302. [PubMed: 15238430]
- Miall RC, Wolpert DM. Forward models for physiological motor control. *Neural Networks*. 1996; 9:1265–1279. [PubMed: 12662535]
- Nolan H, Whelan R, Reilly RB. FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*. 2010; 192:152–162. [PubMed: 20654646]
- Ozaki I, Suzuki Y, Jin CY, Baba M, Matsunaga M, Hashimoto I. Dynamic movement of N100m dipoles in evoked magnetic field reflects sequential activation of isofrequency bands in human auditory cortex. *Clinical Neurophysiology*. 2003; 114:1681–1688. [PubMed: 12948797]
- Pantev C, Eulitz C, Hampson S, Ross B, Roberts LE. The auditory evoked “off” response: sources and comparison with the “on” and the “sustained” responses. *Ear and Hearing*. 1996; 17:255–265. [PubMed: 8807267]
- Poulet JF, Hedwig B. The cellular basis of a corollary discharge. *Science*. 2006; 311:518–522. [PubMed: 16439660]
- Price CJ, Crinion JT, MacSweeney M. A generative model of speech production in Broca's and Wernicke's areas. *Frontiers in psychology*. 2011; 2:1–9. [PubMed: 21713130]
- Reite M, Adams M, Simon J, Teale P, Sheeder J, Richardson D, Grabbe R. Auditory M100 component 1: relationship to Heschl's gyri. *Brain Research: Cognitive Brain Research*. 1994; 2:13–20. [PubMed: 7812174]
- Ross B, Tremblay K. Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Research*. 2009; 248:48–59. [PubMed: 19110047]
- Sams M, Hamalainen M, Antervo A, Kaukoranta E, Reinikainen K, Hari R. Cerebral neuromagnetic responses evoked by short auditory stimuli. *Electroencephalography and Clinical Neurophysiology*. 1985; 61:254–266. [PubMed: 2411504]
- Shahin AJ, Roberts LE, Miller LM, McDonald KL, Alain C. Sensitivity of EEG and MEG to the N1 and P2 Auditory Evoked Responses Modulated by Spectral Complexity of Sounds. *Brain Topography*. 2007; 20:55–61. [PubMed: 17899352]
- Suga N, Shimozawa T. Site of neural attenuation of responses to self-vocalized sounds in echolocating bats. *Science*. 1974; 183:1211–1213. [PubMed: 4812353]
- Suppes P, Han B, Lu Z-L. Brain-wave recognition of sentences. *Proc Natl Acad Sci U S A*. 1998; 95:15861–15866. [PubMed: 9861061]
- Tian X, Poeppel D. The Effect of Imagination on Stimulation: The Functional Specificity of Efference Copies in Speech Processing. *Journal of Cognitive Neuroscience*. 2013; 25:1020–1036. [PubMed: 23469885]
- van der Horst R, Leeuw AR, Dreschler WA. Importance of temporal-envelope cues in consonant recognition. *J. Acoust. Soc. Am.* 1999; 105:1801–1809. [PubMed: 10089603]
- Ventura MI, Nagarajan SS, Houde JF. Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience*. 2009; 10:1–11. [PubMed: 19126204]
- Wager TD, Keller MC, Lacey SC, Jonides J. Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage*. 2005; 26:99–113. [PubMed: 15862210]
- Yuval-Greenberg S, Tomer O, Keren AS, Nelken I, Deouell LY. Transient Induced Gamma-Band Response in EEG as a Manifestation of Miniature Saccades. *Neuron*. 2008; 58:429–441. [PubMed: 18466752]

- Zheng ZZ, Munhall KG, Johnsrude IS. Functional Overlap between Regions Involved in Speech Perception and in Monitoring One's Own Voice during Speech Production. *Journal of Cognitive Neuroscience*. 2010; 22:1770–1781. [PubMed: 19642886]
- Zheng ZZ, Vicente-Grabovetsky A, MacDonald EN, Munhall KG, Cusack R, Johnsrude IS. Multivoxel Patterns Reveal Functionally Differentiated Networks Underlying Auditory Feedback Processing of Speech. *The Journal of neuroscience*. 2013; 33:4339–4348. [PubMed: 23467350]

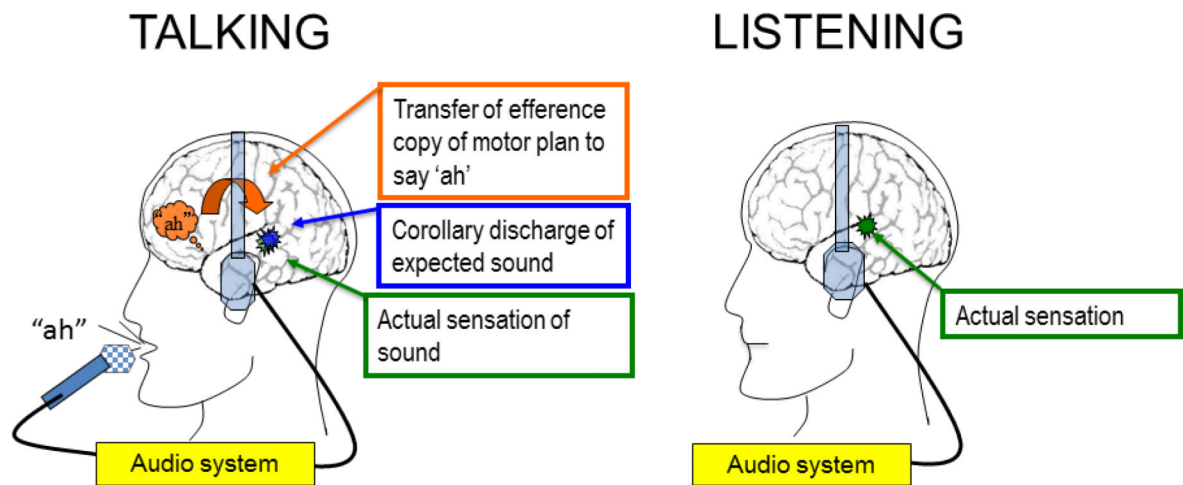


Figure 1.

Illustration of the behavioral tasks. Left shows a cartoon profile of a healthy subject talking (saying “ah”), and right shows listening to a playback of “ah” through headphones. The audio system records the speech sounds during Talking and plays them back during Listening. The intention to say “ah” is indicated as an orange “thought bubble” over the left hemisphere IFG area. The orange curved arrow pointing from the IFG area to auditory cortex indicates the transmission of the efference copy of the motor plan, which produces a corollary discharge (blue burst) of the expected sensation in auditory cortex. When the expected sensation (corollary discharge) matches the actual sensation in auditory cortex (green burst), perception is suppressed.

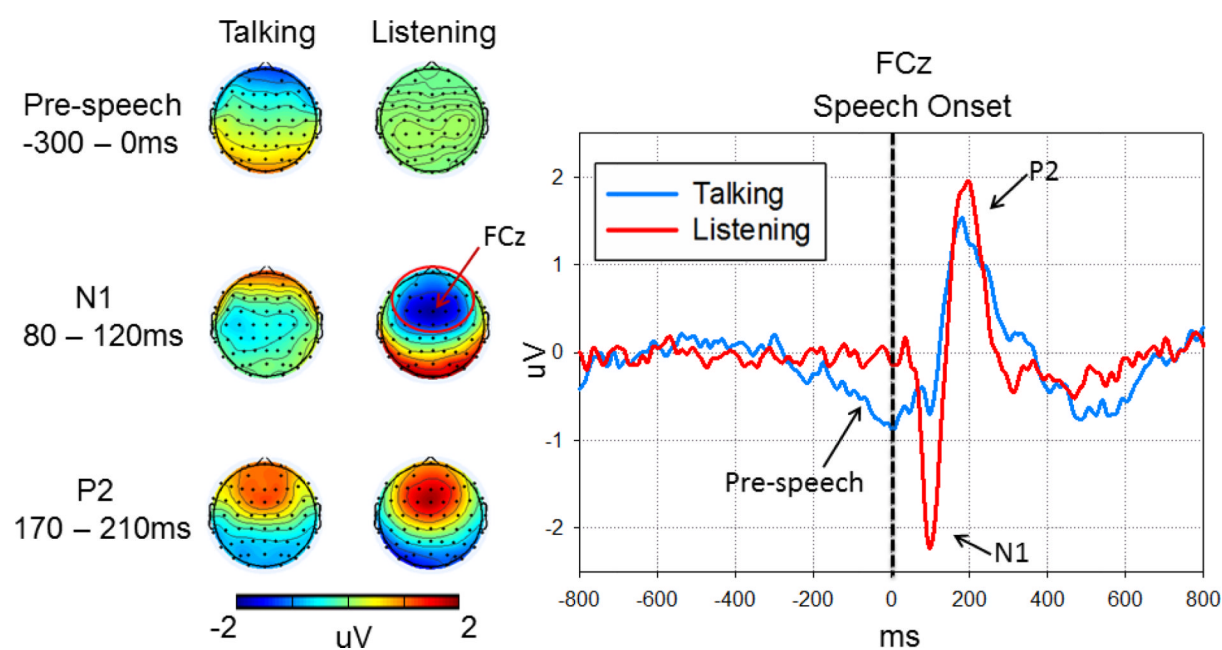


Figure 2.

Mean voltage scalp maps and ERPs in the Talking and Listening tasks. The left voltage scalp plots show spatial distributions of the pre-speech, N1 and P2 ERP responses. Sensors showing maximum activities are marked by red circle surrounding FCz. Note, stronger pre-speech responses were observed in the Talking task while stronger N1 and P2 responses were observed in the Listening task. On the right are ERPs recorded from FCz linked to the onset of the speech sound (dotted vertical line) during both the Listening (red lines) and Talking tasks (blue lines). During Talking, N1 to the speech sound is suppressed relative to N1 to the same sound during Listening. In addition, there is a slow pre-speech negative activity spanned from -300 to 0 milliseconds. Amplitude (microvolts) is on the y-axis and time (milliseconds) is on the x-axis. Vertex negativity is plotted down.

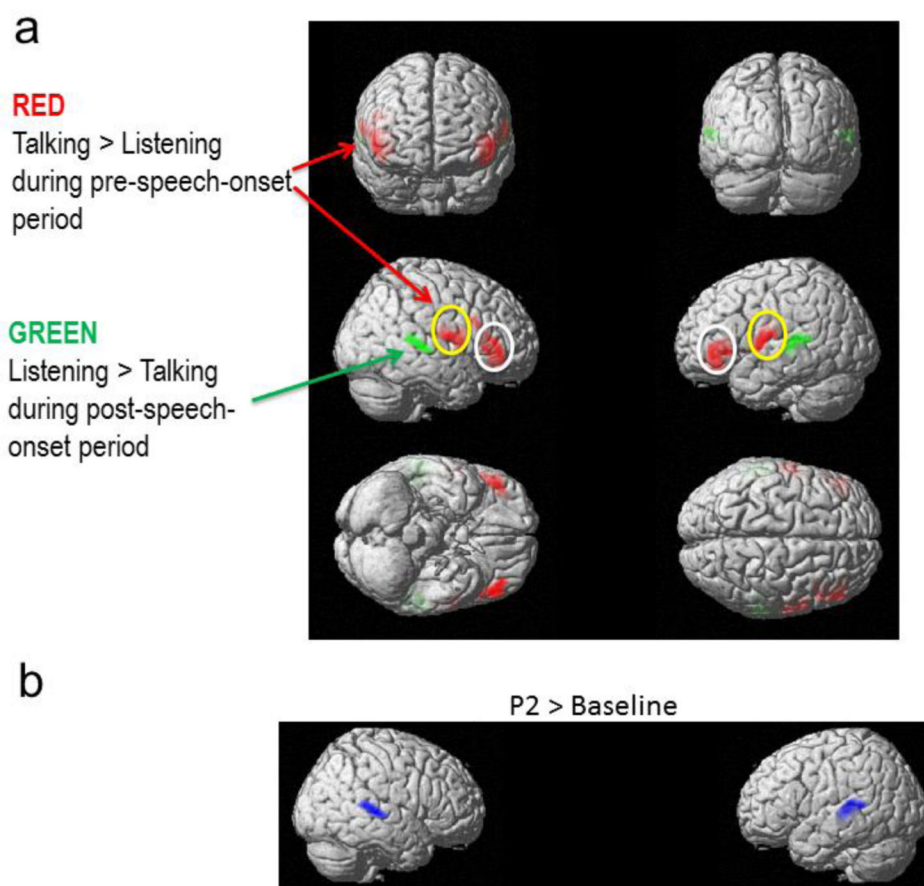


Figure 3.

(a) Source localization maps parsing the Task x Period interactions, showing the Task effect for each time period separately. Red represents greater activity during Talking than Listening, and green represents greater activity during Listening than Talking. Note greater activity in IFG (white circles) and mouth sensorimotor area (yellow circles) during Talking than Listening and greater activity in STG during Listening than Talking. (b) Source localization maps for P2 compared to baseline, averaging across Talking and Listening tasks.

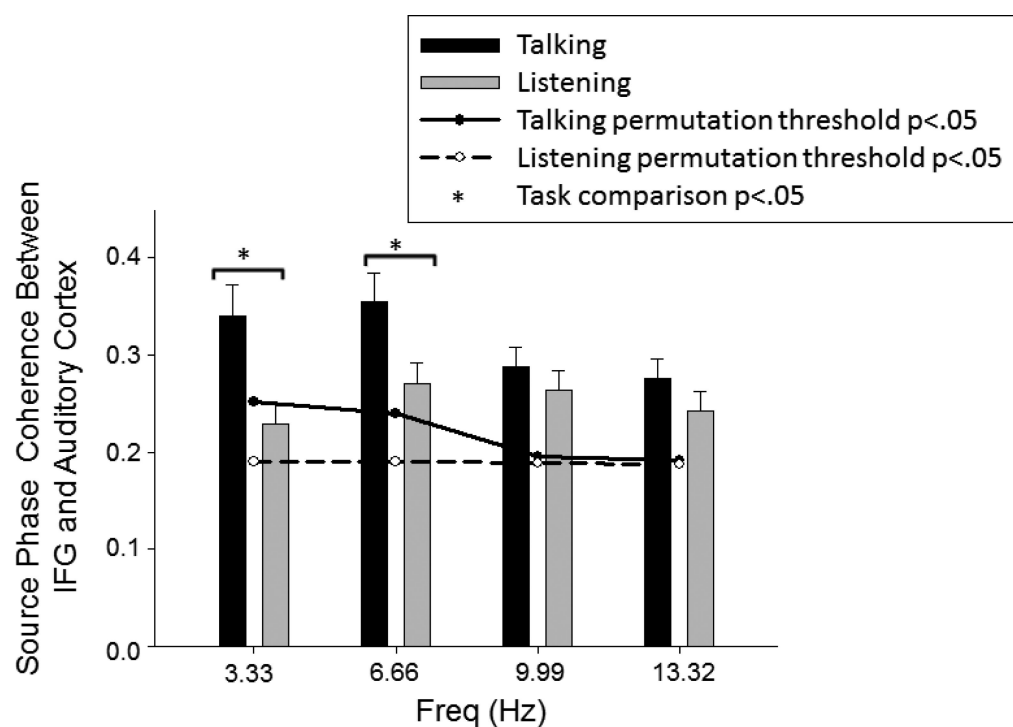


Figure 4.

Frontal-Temporal source coherence for Talking and Listening tasks. The bar graph shows the magnitudes of source coherence during Talking (black bar) and Listening (gray bar) between IFG areas and auditory cortex. Error bars indicate standard error, *significance at $p < 0.05$. Significant thresholds ($p < 0.05$) are based on the results of the permutation analyses for Talking (solid line) and Listening (dash line).

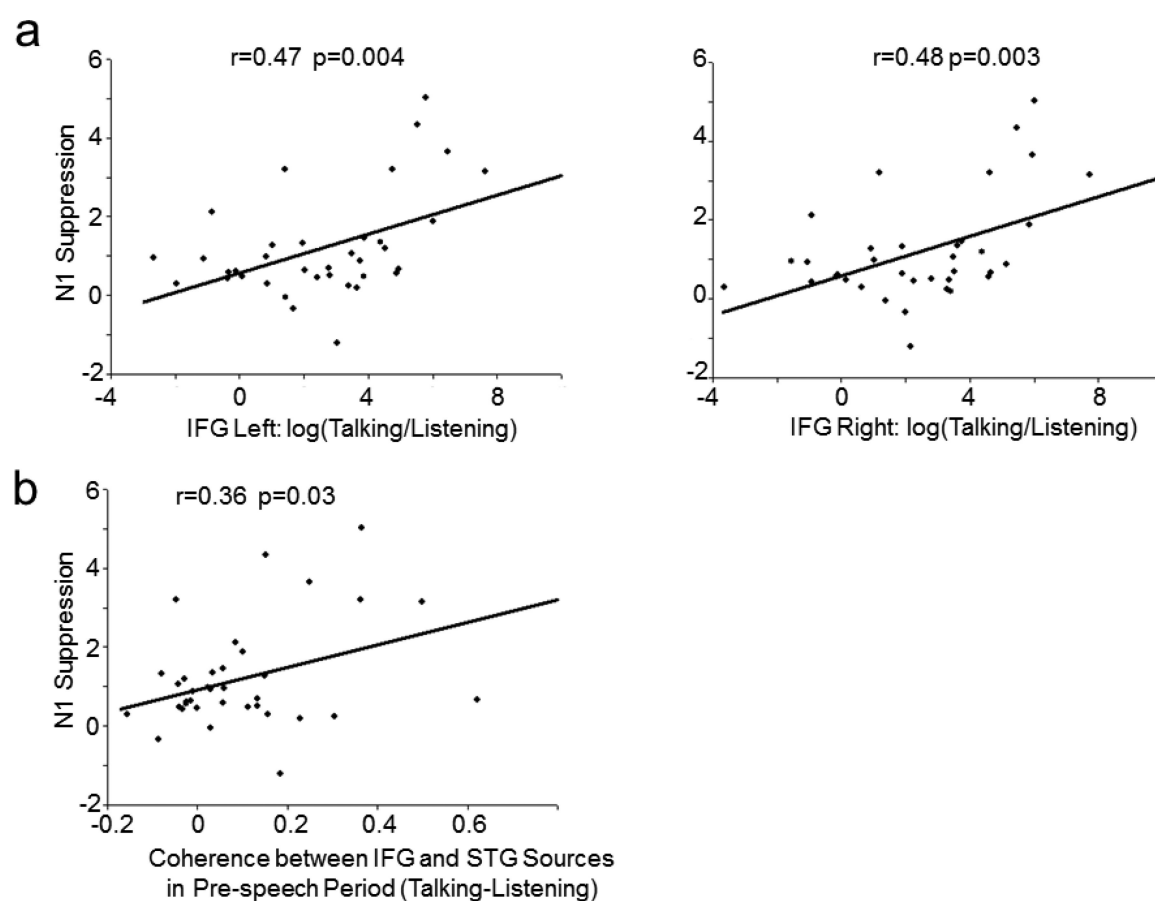


Figure 5.

Bivariate scatter plots depict the relationship between (a) N1 suppression and source activity differences between Talking and Listening in IFG areas, and, (b) N1 suppression and source coherence between IFG area and primary auditory cortex. N1 suppression = N1 (Talking) – N1 (Listening). Source coherence difference = Source coherence (Talking) – Source Coherence (Listening).