HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2014 September 06.

Published in final edited form as:

Nature. 2014 March 6; 507(7490): 94–98. doi:10.1038/nature12935.

Sensory-motor transformations for speech occur bilaterally

Gregory B. Cogan[1], **Thomas Thesen**[2], **Chad Carlson**[2], **Werner Doyle**[3], **Orrin Devinsky**[2,3], and **Bijan Pesaran**[1,*]

[1]Center for Neural Science, New York University, New York, NY 10003

[2]Department of Neurology, NYU School of Medicine, New York, NY 10016

[3]Department of Neurosurgery, NYU School of Medicine, New York, NY 10016

Abstract

Historically, the study of speech processing has emphasized a strong link between auditory perceptual input and motor production output[1–4]. A kind of 'parity' is essential, as both perception- and production-based representations must form a unified interface to facilitate access to higher order language processes such as syntax and semantics, believed to be computed in the dominant, typically left hemisphere[5,6]. While various theories have been proposed to unite perception and production[2,7], the underlying neural mechanisms are unclear. Early models of speech and language processing proposed that perceptual processing occurred in the left posterior superior temporal gyrus (Wernicke's area) and motor production processes occurred in the left inferior frontal gyrus (Broca's area)[8,9]. Sensory activity was proposed to link to production activity via connecting fiber tracts, forming the left lateralized speech sensory-motor system[10]. While recent evidence indicates that speech perception occurs bilaterally[11–13], prevailing models maintain that the speech sensory-motor system is left lateralized[11,14–18] and facilitates the transformation from sensory-based auditory representations to motor-based production representations[11,15,16]. Evidence for the lateralized computation of sensory-motor speech transformations is, however, indirect and primarily comes from lesion patients with speech repetition deficits (conduction aphasia) and studies using covert speech and hemodynamic functional imaging[16,19]. Whether the speech sensory-motor system is lateralized like higher order language processes, or bilateral, like speech perception is controversial. Here, using direct neural recordings in subjects performing sensory-motor tasks involving overt speech production, we show that sensory-motor transformations occur bilaterally. We demonstrate that electrodes over bilateral inferior frontal, inferior parietal, superior temporal, premotor, and somatosensory cortices exhibit robust sensory-motor neural responses during both perception and production in an overt

Correspondence should be addressed to: Bijan Pesaran, Ph.D., 4 Washington Pl. Rm 809, New York, NY 10003, Tel: 212.998.3578, Fax 212.995.4011, bijan@nyu.edu.

Author Contributions: GC designed the experiment, performed the research, analyzed the data, and wrote the manuscript. TT and OD performed the research and wrote the manuscript. CC and WD performed the research. BP designed the experiment, performed the research, analyzed the data, and wrote the manuscript.

Competing Financial Interests: There are no competing financial interests.

Supplementary Information: Supplementary Information contains Supplementary Discussion, Supplementary Tables, Supplementary Figures S1–S11, and Supplementary References.

word repetition task. Using a non-word transformation task, we show that bilateral sensory-motor responses can perform transformations between speech perception- and production-based representations. These results establish a bilateral sublexical speech sensory-motor system.

To investigate the sensory-motor representations that link speech perception and production, we used electrocorticography (ECoG), in which electrical recordings of neural activity are made directly from the cortical surface in a group of patients with pharmacologically-intractable epilepsy. ECoG is an important electrophysiological signal recording modality that combines excellent temporal resolution with good spatial localization. Critically for this study, ECoG data contain limited artifacts due to muscle/movements during speech production compared with non-invasive methods which suffer artifacts with jaw movement[20]. Thus, using ECoG we were able to directly investigate neural representations for sensory-motor transformations using overt speech production.

Sixteen patients with subdural electrodes (see Fig S1, S2) implanted in the left hemisphere (6 subjects), right hemisphere (7 subjects), or both hemispheres (3 subjects) performed variants of an overt word repetition task designed to elicit sensory-motor activations (Fig 1A, **Online Methods,** and Table S1). We observed increases in neural activity across the high gamma frequency range (60 – 200 Hz and above) with maximal activity across subjects between 70–90 Hz. High gamma activity reflects the spiking activity of populations of neurons during task performance[20,21]. Individual electrodes showed one of three types of task responses: Sensory-motor (S-M), Production (PROD) or Auditory (AUD; Fig 1b, see **Online Methods**). We found that AUD activity was generally localized to the superior temporal gyrus and middle temporal gyrus (42/57 electrodes – 74% – green; Fig 2a, b) and PROD activity occurred mostly in the pre/motor cortex, somatosensory cortex, and the inferior parietal lobule (98/124 electrodes – 79% – blue; Fig 2a, b), in keeping with previous models/results of speech perception and production[11,12,17]. Furthermore, electrical stimulation of PROD electrode locations resulted in orofacial movements consistent with a motor function (see Fig S3). Critically, contrary to one of the core dogmas of brain and language, S-M activity occurred bilaterally in the supramarginal gyrus, middle temporal gyrus, superior temporal gyrus, somatosensory cortex, motor cortex, premotor cortex and inferior frontal gyrus (red; Fig 2a, b, 49 electrodes, see Table S2, Fig S4) and was observed in all subjects (Fig 2a). Of the 49 S-M sites, 45 sites showed auditory activation during the Listen task (red with green outlines; Fig 2a, b; Fig S4,S5; 45/49 ~ 92 %), suggesting a role in speech perception. Hemispheric dominance as determined by Wada testing did not correlate with the hemisphere of the electrode placement ($\chi^2(3) = 0.92$, $p = 0.34$). Importantly, in three subjects with bilateral coverage, S-M activity was present on electrodes in both hemispheres (Fig 2a, c) and the likelihood of an electrode being a S-M site did not differ between hemispheres (Fisher's exact test, $p = 0.31$). These results demonstrate that S-M activity occurs bilaterally.

Given the evidence for bilateral S-M activity, we performed a series of analyses and experimental manipulations to test the hypothesis that bilateral S-M activity is in fact sensory-motor and represents sensory-motor transformations for speech.

One concern is that S-M activity is not due to sensory and motor processes but rather to sensory activation in both auditory (input) and production epochs (sound of your own voice). We observed several convergent lines of evidence that S-M activity reflects both sensory and motor processing (see Fig 2d and **Online Methods**). **1)** S-M sites contain a sensory response because they responded to auditory stimulation as rapidly as AUD sites. (S-M latency = 158 ms; AUD = 164 ms, see Fig 2d). **2)** S-M responses during production are not due to auditory sensory reafferent input from hearing one's own voice because responses were present during the Listen-Mime task as well as the Listen-Speak task. **3)** S-M responses during production are not due to somatosensory reafference from moving articulators because S-M activity significantly increased within 248 ms of the production Go cue while vocal responses occurred substantially later at 1,002 ms (+/− 40 ms SEM). **4)** S-M production responses contain motor features because they occurred together with, and even before, PROD electrode production responses, (S-M = 248 ms; PROD = 302 ms, q = 0.03; Permutation test see **Online Methods**). **5)** S-M activity was persistently elevated during the delay period (p= 0.01; see Fig 2d, **Online Methods**), broadly consistent with planning activity, unlike PROD delay period activity (p = 0.64) or AUD delay period activity (p = 0.53). These results demonstrate that S-M activity cannot be simply sensory and spans both sensory and motor processes.

A related concern is that sensory-motor transformations are first carried out in the left hemisphere (LH). If so, S-M responses in the right hemisphere (RH) could be due to communication from the LH. To test this hypothesis, we further examined latencies of S-M responses according to hemisphere. Response latencies did not differ significantly in each hemisphere in either the auditory (RH: 156 ms; LH: 182 ms; $q = 1 \times 10^{-4}$; Permutation test) or the production epoch (RH: 272 ms; LH: 268 ms; $q = 1 \times 10^{-4}$ – see **Online Methods**). Therefore, RH responses cannot be due to computations that were first carried out in the LH and the data do not support strictly lateralized sensory-motor computations.

Another concern is that S-M activity may not reflect speech processing and may also be present during simple auditory inputs and orofacial motor outputs. To test this, we employed a Tone-Move task in one of the bilaterally implanted subjects (S13; see **Online Methods**). We found that S-M electrodes did not have significant sensory-motor responses during the Tone-Move task (p = 0.36; permutation test; see Fig S6). Thus, S-M activity is specific to mapping sounds to goal-directed vocal motor acts and is likely specific to speech (see Supplementary Discussion 1.3).

Thus far, we have shown the S-M activity is bilateral, sensory-motor, and likely specific to speech. However, an important open question is whether S-M responses reflect the transformation that links speech perception and production and can support a unified perception-production representational interface. A specific concern is that high gamma ECoG activity may pool heterogeneous neural responses beneath the electrode. S-M responses may combine activity from neurons which encode perceptual processes active during the auditory cue and other neurons which encode production processes active during the utterance. If this is true, none of the activity necessarily reflects a sensory-motor transformation that links perception and production. To be able to rule out this alternative and demonstrate that S-M responses are involved in sensory-motor transformations, we

reasoned that two requirements must be met. S-M activity must encode information about the content of the underlying speech processes, and this encoding must reflect transformative coding between the sensory input and motor output.

To test whether S-M activity encodes information about speech content, we decoded the neural activity to predict, on each trial, what the subjects heard and said. We used the seven consonant-vowel-consonant words and trained a 7-way linear classifier to decode the neural responses (see **Online Methods**). Individual electrodes only weakly encoded speech content, but when we decoded activity pooled across groups of electrodes, we found that all three electrode groups encoded speech tokens (see Fig 3). AUD electrodes performed best with an average classification performance of 42.7% (χ^2 (1) = 56.5, p = 6×10^{-14}), followed by S-M electrodes - 33.4% (χ^2 (1) = 25.6, p = 4×10^{-7}), and then PROD electrodes - 27.1% (χ^2 (1) = 11.5, p = 7×10^{-4}). Furthermore, classification performance for S-M electrodes did not differ between the two hemispheres (LH: 29%. RH: 27%. Fisher's exact test, p = 0.5, Fig 3c). Thus, bilateral S-M activity encodes information about the sensory and motor contents of speech, meeting an important requirement for sensory-motor transformations.

We next sought to test whether S-M activity can link speech perception and production by transforming auditory input into production output. The essential requirement for transformation is that neural encoding of sensory input should depend on subsequent motor output. Previous work has characterized visual-motor transformations using a transformation task in which the spatial location of a visual cue can instruct a motor response to the same or different spatial location – the pro-anti task[22,23]. Sensory-motor neurons in the dorsal visual stream display different responses to the visual cue depending on the motor contingency, demonstrating a role for these neurons in the visual-motor transformation[22].

Given these predictions from animal neurophysiology, we tested four subjects as they performed an auditory-motor transformation task - the Listen-Speak Transformation task – that employed two non-words (kig, pob) to examine whether S-M activity plays a role in transformations for speech (see Fig 4a, Fig S7, Fig S8 and **Online Methods)**. This task allowed us to hold the sensory and motor components constant while manipulating the transformation process itself in order to measure how the encoding of this content changed depending on how perceptual input was mapped onto production output. The use of non-words instead of words offered other advantages. Non-words allowed us to examine sublexical transformations for speech and could be designed to differ maximally in their articulatory dimensions and their neural representations (see **Online Methods and** Supplementary Discussion 1.1, 1.2).

At least three models describe how neural responses encode the task variables. If responses follow a strictly Sensory model, the encoding will follow the content of the sensory inputs and confuse **kig**→kig with **kig**→pob trials and **pob**→pob with **pob**→kig trials (see Fig 4b.i). Conversely, responses that follow a strictly Motor model will encode the production outputs, confusing kig→**kig** with pob→**kig** trials and pob→**pob** with kig→**pob** trials (see Fig 4b.ii). If S-M responses pool responses from sensory and motor neurons, the encoding will follow the sensory model during sensory input and the motor model during motor output. In contrast, S-M responses that reflect the transformation of sensory input into motor

output must follow a different Transformation model and encode the sensory information differently depending on the upcoming motor act (see Fig 4b.iii). Neural activity displaying this property could compute a representational transformation (see Supplementary Discussion 1.1, 1.2). If so, responses that follow a Transformation model will not confuse trial conditions with either identical input or identical output. Consequently, each of the three models predicted very different patterns of neural coding.

We constructed linear classifiers to decode neural responses. As expected, AUD electrodes in the auditory epoch encoded the auditory input (Fig 4b.i, Fig 4c.i) and PROD electrodes encoded the output during the production epoch (utterance – Fig 4b.ii, Fig 4c.ii). S-M electrodes however, changed their encoding over the course of the trial. During the auditory epoch, S-M electrodes encoded both sensory and motor conditions concurrently, consistent with the presence of a sensory-motor transformation (Fig 4b.iii, Fig 4c.iii). Interestingly, during the production epoch, S-M responses no longer encoded the auditory input and encoded the production output (Fig 4c.iv) suggesting the transformation has largely been computed by that time. To quantify the comparison of different models, we used the Kullback-Leibler (K-L) divergence (see Fig 4d.i–iv, **Online Methods**). The results were consistent with the response patterns in the confusion matrices.

We can also rule out that the difference in S-M responses is due to a third population of neurons that selectively responds to the cue instructing how perceptual input was mapped onto production output (‘match’ or ‘mismatch’). We ran the same linear classifier during cue presentation and found that the S-M electrodes did not encode the cue (χ^2 (1) = 0.08, p = 0.78, see **Online Methods**).

Using direct brain recordings (ECoG) and overt speech, we demonstrate that a sensory-motor system for transforming sublexical speech signals exists bilaterally. Our results are in keeping with models of speech perception that posit bilateral processing but contradict models that posit lateralized sensory-motor transformations[11,16]. Our results also highlight how S-M activity during perceptual input reflects the transformation of speech sensory input into motor output. We propose that the presence of such transformative activity demonstrates a unified sensory-motor representational interface that links speech perception- and production-based representations. Such an interface is important during speech articulation, acquisition, and self-monitoring[24–26]. Since right hemisphere lesions do not give rise to conduction aphasia[19,27–29], our evidence for bilateral sensory-motor transformations promotes an interesting distinction between speech and language: While sensory-motor transformations are bilateral, the computational system for higher order language is lateralized[5,6] (see Fig S9). This hypothesis invokes a strong interface between sensory-based speech perception representations and motor-based speech production representations and suggests that deficits for conduction aphasia are more abstract/linguistic in nature. We propose that bilateral sublexical sensory-motor transformations could support a unification of perception- and production-based representations into a sensory-motor interface[6], drawing a distinction between the bilateral perception-production functions of speech and lateralized higher order language processes.

Online Methods

Participants

Electrocorticographic (ECoG) recordings were obtained from 16 patients (7 males, 9 females - Table S1) with pharmacologically-resistant epilepsy undergoing clinically motivated subdural electrode recordings at the NYU School of Medicine Comprehensive Epilepsy Center. Informed Consent was obtained from each patient in accordance with the Institutional Review Board at NYU Langone Medical Center. Patient selection for the present study followed strict criteria: 1) Cognitive and language abilities in the average range or above, including language and reading ability, as indicated by formal neuropsychological testing (see Table S1); 2) Normal language organization as indicated by cortical stimulation mapping, when available. In addition, only electrode contacts outside the seizure onset zone and with normal inter-ictal activity were included in the analysis.

Surface reconstruction and electrode localization

To localize electrode recording sites, pre-surgical and post-surgical T1-weighted MRIs were obtained for each patient and co-registered with each other[31]. The co-registered images were then normalized to an MNI-152 template and electrode locations were then extracted in MNI space (projected to the surface) using the co-registered image, followed by skull-stripping[32]. A three-dimensional reconstruction of each patient's brain was computed using FreeSurfer to generate Fig 2, S2, S3, S4, S5, S6, S7, S8, and S10[33]. For Table S2, Talairach coordinates were converted from MNI space using the EEG/MRI toolbox in Matlab (http://sourceforge.net/projects/eeg/, GNU General Public License).

Behavioral tasks and recordings

All participants performed three behavioral tasks: Listen-Speak, Listen-Mime and Listen (Fig 1a). Behavioral tasks were performed while participants were reclined in a hospital bed. Tasks were controlled by a PC placed on the service tray on the bed running the Presentation program (NeuroBehavioral Systems). Behavioral audio recordings were either synchronized with the neural recordings at 10 kHz (see below) or recorded on the PC and referenced to the 'Go' Cue. For a subset of subjects, a video camera with built-in microphone (Sony) was positioned to monitor subject orofacial movements and utterances. Video was streamed to disk (Adobe Premier Pro. Video: 29.95 fps. Audio: 44.1 kHz). Audio-visual and neural signals were synchronized video frame-by-video frame using an Analog-to-Digital Video Converter (Canopus).

Listen-Speak, Listen-Mime, and Listen tasks were randomly interleaved on a trial-by-trial basis with at least 4 s between trials. Each trial began with a visual cue presented, followed by the auditory consonant-vowel-consonant (CVC) token 500 ms later. We used CVC words composed of the same consonants, 'h' and 't', and different vowels – hat, hit, heat, hoot, het, hut, hot. These spoken syllables span the vowel space and differ in their auditory and articulatory content. Subjects had to either listen passively ('Listen'), repeat the syllable after a cue ('Listen-Speak'), or mime the syllable after a different cue (produce the appropriate mouth movements but with no vocal cord vibration – 'Listen-Mime'[30]; see Fig S10). The temporal delay between the auditory cue and the movement cue was 2 s. We

obtained between 49 and 166 trials per condition (within subject) and between 175 and 334 total trials per subject.

For the Tone-Move task (see Fig S6), after the 'Listen' cue was delivered, a 500 ms, 1000 Hz sinusoidal tone (with 100 ms on and off ramps) was presented. After a short, 2 s delay another visual cue was presented ('Move') instructing the subject to move their articulators (tongue, lips, and jaw). For one subject, these trials were randomly interleaved within blocks of the Listen-Speak/Listen-Mime/Listen tasks (see above).

For the Listen-Speak Transformation task, four subjects (see Fig S7, S8) were first presented with one of two visual cues: 'Match - Listen or 'Mismatch - Listen'. After a delay, subjects heard one of two non-words: 'kig' (/kIg/) or 'pob' (/pɔb/). These non-words were chosen to differ maximally on their articulator dimensions: 'kig' contains a velar (back) voiceless stop consonant, followed by a high front vowel and finally a velar voiced stop consonant, while 'pob' contains a bilabial (front) voiceless stop consonant followed by a back low vowel and then a bilabial front voiced stop consonant. The tongue movement therefore goes back to front to back for 'kig' and front to back to front for 'pob'. The reason for choosing maximally different articulations was that larger articulator differences might lead to larger neural activity differences. After a short delay (randomized between 1.5 and 2 s), another visual cue was presented ('speak') at which time subjects were to either say the match non-word they had heard if they had seen the initial match cue, or say the mismatch non-word if they had seen the mismatch cue. Each token within each condition was presented between 63 and 78 times per subject, with total trials ranging from 255 to 309 per subject. This control was carried out in separate blocks trials that alternated with blocks of the main Listen-Speak/Listen-Mime/Listen task.

Neural recordings and preprocessing

EEG data were recorded from intracranially implanted subdural electrodes (AdTech Medical Instrument Corp., WI, USA) in patients undergoing elective monitoring of pharmacologically intractable seizures. Electrode placement was based entirely on clinical grounds for identification of seizure foci and eloquent cortex during stimulation mapping, and included grid (8×8 contacts), depth (1×8 contacts) and strip (1×4 to 1×12 contacts) electrode arrays with 10 mm inter-electrode spacing center-to-center. Subdural stainless steel recording grid and strip contacts were 4 mm in diameter; consequently the distance between contacts was 6mm and they had an exposed 2.3 mm diameter recording contact.

For seven of the 16 subjects, neural signals from up to 256 channels were amplified (10 x, INA121 Burr-Brown instrumentation amplifier), bandpass filtered between 0.1 – 4000 Hz and digitized at 10 kHz (NSpike, Harvard Instrumentation Labs) before being continuously streamed to disk for off-line analysis (custom C and Matlab code). The front-end amplifier system was powered by sealed lead acid batteries (Powersonic) and optically isolated from the subject. After acquisition, neuronal recordings were further low-pass filtered at 800 Hz and down-sampled offline to 2000 Hz for all subsequent analysis. For the remaining nine subjects, neural signals from up to 128 channels were recorded on a Nicolet One EEG system, bandpass filtered between 0.5 – 250 Hz and digitized at 512 Hz. In some recordings, modest electrical noise was removed using line-filters centered on 60, 120, and 180 Hz[34].

Data Analysis

Activation Analysis—Time-frequency decomposition was performed using a multi-taper spectral analysis[34]. The power spectrum was calculated on a 500 ms analysis window with ±5 Hz frequency smoothing stepped 50 ms between estimates. Single trials were removed from the analysis if the raw voltage exceeded eight standard deviations from the across trial pool, and noisy channels were removed from the analysis by visual inspection or if they did not contain at least 60 % of the total trials after the standard deviation threshold removal.

Sensory-motor transformations were defined as activity in the gamma range (70–90 Hz) that followed the auditory stimulus as well as the production cue during both Listen-Speak and Listen-Mime (Fig 1b). As the example responses illustrate, some electrodes showed consistent increases in activity in the high gamma band as high as 300 Hz. Since the frequency extent varied across subjects, we chose to focus on the 70–90 Hz frequency range as this band showed the greatest activation consistently across all subjects. Similar results were obtained when a broader frequency range extending up to 150 Hz was analyzed. While the Listen-Mime condition does involve altering the motor plan (no vocal cord vibration), sensory-motor activations were based on the conjunction of activity in both the Listen-Speak and the Listen-Mime conditions. Any neural activity that was specific to the Listen-Mime condition and not present in 'normal' speaking conditions was therefore excluded (see Fig S10).

Responses were divided into three types. The first response type, Auditory (AUD), was defined as containing a response that was seen within 250–750 milliseconds following the onset of the auditory stimulus in all three conditions (Fig 1b *top*). The second response type, Production (PROD), was characterized as containing a response occurring between 500–1000 milliseconds after the respond cue in the Listen-Speak and the Listen-Mime conditions (Fig 1b **middle**). The last response type, sensory motor (S-M), contained both post stimulus and a post response cue activation in both the Listen-Speak and the Listen-Mime conditions (Fig 1b *bottom*). The baseline period was defined as the 500 ms preceding the auditory stimulus.

In Figure 1b, the experimental epoch was defined as −500 ms to 3500 ms post auditory stimulus onset. In Figure 2c the experimental epoch was defined as −500 ms to 4000 ms post auditory stimulus onset. The additional 500 ms was included in Figure 2c to compensate for slightly later production responses for that the subject. Power in each frequency band was normalized to the power in the baseline period by dividing by the power at each frequency. Since the neural responses had variable onset times but were on average quite long in duration, the times were chosen to adequately sample all the responses under investigation.

To assess statistical significance, the average power across trials was taken in two time regions of interest for each trial within each condition. For the Listen condition, the baseline values for each trial were shuffled with the post auditory values 10,000 times to create a null distribution. For the Listen-Speak and the Listen-Mime conditions, both the post-auditory and the post-production values were shuffled 10,000 times with the baseline values to create two null distributions. Initial significance was assessed using a permutation test by comparing the actual difference between the post auditory and post production values with

the shuffled data differences[35]. To correct for multiple comparisons, for all subjects, all three conditions and both analysis epochs (Listen – post auditory, Listen-Speak – post auditory and post production, and Listen-Mime – post auditory and post production) were pooled together and a false discovery rate (FDR) analysis was performed with an alpha threshold set at 0.05[36].

Population average latency analysis—The population latency analysis was performed using the baseline corrected high gamma power response profiles for each electrode within each response class (S-M, AUD, PROD). The high gamma neural responses were first band-pass filtered (70–90 Hz) and then averaged within conditions. The Listen-Speak and Listen-Mime conditions were averaged together. Since the data were recorded using two different sampling rates, the data were resampled to a 500 Hz sampling rate. To test for latencies within a response class, the latencies following either the auditory onset or the go cue were compared against the activity in the Listen condition following the go cue by computing a permuted distribution for each time point. The significance values at each time point were then corrected for multiple comparisons using a false discovery rate set with an alpha of 0.05. The first time point that was followed by at least 20 consecutive significant time points (40 ms) was taken to be the latency of the neural response. This resulted in four significant latency values. In the sensory epoch, AUD electrodes had significant neural responses at 164 ms and S-M electrodes had significant responses at 158 ms. During the motor epoch, PROD electrodes had significant responses starting at 302 ms, while S-M electrodes had significant responses starting at 248 ms. A similar analysis was carried out comparing the left S-M electrodes with the right S-M electrodes, which resulted in four more significant latency values: RH-sensory 156 ms, LH-sensory 182 ms, RH-motor 272 ms, and LH-motor 268 ms. A direct comparison between these latencies within each task epoch using FDR-corrected shuffle tests (see above) revealed no significant results.

To assess whether or not during the sensory and motor epochs, the S-M electrodes display significantly faster neural responses than the AUD/PROD electrodes, we repeated the permutation test, except instead of using the comparison of the task compared to the ‘Listen’ Condition, we compared the S-M electrodes to the AUD electrodes in the sensory epoch and the S-M electrodes compared the PROD electrodes in the motor epoch. The results showed that while S-M and AUD electrodes did not differ in their latency values during the sensory epoch, S-M electrodes were significantly faster than PROD electrodes in the motor epoch.

Power Analysis—To test for power differences of the high gamma response (70–90 Hz) across hemispheres, we performed FDR-corrected permutation tests. Data were analyzed by averaging a 300 ms time window, sliding 50 ms between estimates. The data were baseline corrected (average – 500 ms – 0 ms pre-stimulus activity across conditions, within electrodes) and then log transformed prior to analysis. For each condition – Listen-Speak, Listen-Mime, and Listen and within each hemisphere (left and right), we computed the task epoch responses by computing the average of the high gamma response during the auditory epoch (0 – 1000 ms post auditory onset) and during the production epoch (0 – 1500 ms post production cue onset). We then performed a series of permutation tests where we permuted the neural response across condition and/or across hemisphere, correcting for multiple

comparisons using a FDR procedure. Only four tests produced significant results: Listen-Speak vs. Listen during the production epoch in each hemisphere, and Listen-Mime vs Listen during the production epoch in each hemisphere. Furthermore, the neural responses within all conditions was not different across hemispheres (see Fig S11, $p > 0.05$, FDR corrected).

Delay activity Analysis—To assess the significant delay activation for each electrode class, a permutation test was carried out using filtered data as listed above. A permutation test was performed for each electrode class in which the average high gamma neural activity of the delay period (1 – 2 s post auditory onset) was compared to that of the baseline period (−1 s to −0.5 s pre auditory onset). While PROD electrodes and AUD electrodes did not display elevated population neural activity (p = 0.64 and 0.53 respectively), S-M electrodes had significantly higher elevated delay activity compared to baseline (p = 0.02; see Fig 2d).

Classifier—Classification was performed using the single value decomposition (SVD) of the high gamma neural response (70 – 160 Hz – 300 ms sliding windows with an overlap of 250 ms) in either the auditory epoch (0–1000 ms post Auditory Onset – AUD Electrodes) or the production epoch (0–1500 ms post Go Cue – PROD Electrodes), or both (S-M Electrodes). A linear discriminant analysis (LDA) classification was performed using a leave-one-out validation method, where the training set consisted of all the trials of the data set except the one being tested. Note that analyzing the different task epochs separately for the S-M electrodes produced classifier results that were also significantly above chance (auditory epoch: 40.2 % - (χ^2 (1) = 47, p = 7×10^{-12}, production epoch: 23.2 % - χ^2 (1) = 5.6, p = 0.02).

To create the cumulative curves, the number of electrodes inputted into the classifier was increased linearly. To control for the variability in trial numbers, the minimum number of trials common to all subjects/electrodes was used. 100 iterations for each number of cumulative electrodes were performed where the specific trials and the specific electrodes were varied randomly, while the number of SVD components was equal to the number of electrodes inputted to the classifier for the AUD and S-M electrodes, whereas five components were used for the PROD electrodes due to a lower number of components present in the PROD electrode data.

Confusion matrix scores are simply the proportion of trails classified as the token on the horizontal axis (decoded) given that the actual trial came from the vertical axis (actual). Confusion matrices in Fig 3a are shown for the largest number of cumulative electrodes in each electrode class.

To analyze the Listen-Speak Transformation task responses (Fig 4), the same decomposition (SVD) of the neural signal (70–160 Hz) was used. Note that instead of a 7-way classifier, a 4-way classifier was used. Confusion matrices (Fig 4c) are shown for the largest number of cumulative electrodes in each electrode class (AUD = 10; PROD = 19; S-M = 8). For the S-M electrodes, each response epoch (auditory – Fig 4c.iii and production – Fig 4c.iv) was analyzed separately.

To measure the quality of each of the models (Sensory, Motor, Sensory-Motor or Chance; Fig 4d) we used the Kullback-Leibler Divergence which quantifies the amount of information lost in bits when Q (the model) is used to approximate P (the data):

$$D_{KL}(P\|Q)=\sum_i P(i)\log_2\left(\frac{P(i)}{Q(i)}\right)$$

Where P is the classification percentage for each actual/decoded pair (see above) and Q is one of the four models: Sensory, Motor, Sensory-motor and Chance. The K-L divergence estimates the information distance between the pattern of classification errors predicted by each model, shown in Fig 4B and the pattern of classification errors based on neural recordings, shown in Fig 4C. Smaller K-L divergence reflects more information about classification errors and improved model fit. The Sensory model (Fig 4b.i) reflects classifications scores that track the auditory speech input such that in both the match and the mismatch cases, the same input will be confused with each other. Conversely, the Motor model (Fig 4b.ii) reflects classification scores that track the production output so that the same outputs will be confused with one another. The Sensory-motor model (Fig 4b.iii) however will reflect both the input and output such classifications for each of the conditions presented (Kig → Kig, Pob → Pob, Kig → Pob, and Pob → Kig) will be classified correctly. Lastly, the Chance model will simply reflect chance performance in all cases (0.25).

Classifier analysis of the cue data (‘Match Listen’ vs. ‘Mismatch Listen’) in the Listen-Speak Transformation task was analyzed on the SM electrodes for the subjects performing the task. The same linear classifier was used as above, but was performed during the cue period (0–1000 ms post Cue) and was 2-way (‘Match Listen’ vs. ‘Mismatch Listen’ Cue). The results demonstrated that the classification was not significant (mean classification = 52.3 %, χ^2 (1) = 0.08, p = 0.78). Furthermore, using the same 2-way classifier between the match and mismatch condition during the auditory epoch was also not significant (mean classification = 56.4 %, χ^2 (1) = 0.72, p = 0.4). Taken together, this indicates that the sensory-motor transformations displayed by these electrodes cannot be due to a third population of neurons that code for the visual cue.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Adam Weiss, Jim MacArthur and Loren Frank for developing the data acquisition hardware and software, Olga Felsovalyi, Elizabeth Londen, Priya Purushothaman, Lucia Melloni, Callah Boomhaur, and Amy Trongnetrpunya for technical assistance, and David Poeppel and Carlos Brody for comments on the manuscript. This work was supported, in part, by R03-DC010475 from the NIDCD, a Career Award in the Biomedical Sciences from the Burroughs Wellcome Fund (BP), a Watson Investigator Program Award from NYSTAR (BP), a McKnight Scholar Award (BP) and a Sloan Research Fellowship (BP).

References

1. Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the Speech Code. Psychol Rev. 1967; 74:431–461. [PubMed: 4170865]
2. Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21:1–36. [PubMed: 4075760]
3. Halle M, Stevens KN. Speech recognition: A model and a program for research. IEEE Trans Information Theory. 1962; 8:155–159.
4. Halle M, Stevens KN. Analysis by synthesis. Proceedings of seminar on speech compression and processing. 1959; D7
5. Berwick RC, Friederici AD, Chomsky N, Bolhuis JJ. Evolution, brain, and the nature of language. Trends Cogn Sci. 2013; 17:89–98. [PubMed: 23313359]
6. Chomsky, N. The minimalist program. MIT Press; Boston, MA: 1995.
7. Jakobson, R. Child Language, Aphasia and Phonological Universals. Mouton; The Hague: 1968.
8. Lichtheim L. On aphasia. Brain. 1885; 7:433–485.
9. Wernicke C. The aphasic symptom-complex: A psychological study on an anatomical basis. Arch Neurol. 1970; 22:280–282.
10. Geschwind N. Disconnexion syndromes in animals and man. I. Brain. 1965; 88:237–94. 585–644. [PubMed: 5318481]
11. Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007; 8:393–402. [PubMed: 17431404]
12. Price CJ. The anatomy of language: a review of 100 fMRI studies published in 2009. Ann N Y Acad Sci. 2010; 1191:62–88. [PubMed: 20392276]
13. Obleser J, Eisner F, Kotz SA. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J Neurosci. 2008; 28:8116–23. [PubMed: 18685036]
14. Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Rev Neurosci. 2009; 12:718–24.
15. Hickok G, Houde J, Rong F. Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. Neuron. 2011; 69:407–422. [PubMed: 21315253]
16. Hickok G, Okada K, Serences JT. Area Spt in the Human Planum Temporale Supports Sensory-Motor Integration for Speech Processing. J Neurophysiol. 2009:2725–2732. [PubMed: 19225172]
17. Guenther FH. Cortical interactions underlying the production of speech sounds. J Commun Disord. 2006; 39:350–65. [PubMed: 16887139]
18. Wise RJS, et al. Separate neural subsystems within “Wernicke”s area’. Brain. 2001; 124:83–95. [PubMed: 11133789]
19. Caramazza A, Basili AG, Koller JJ, Berndt RS. An investigation of repetition and language processing in a case of conduction aphasia. Brain Lang. 1981; 14:235–71. [PubMed: 7306783]
20. Crone NE, Sinai A, Korzeniewska A. Event-Related Dynamics of Brain Oscillations. Prog Brain Res. 2006; 159:275–95. [PubMed: 17071238]
21. Markowitz DA, Wong YT, Gray CM, Pesaran B. Optimizing the decoding of movement goals from local field potentials in macaque cortex. J Neurosci. 2011; 31:18412–22. [PubMed: 22171043]
22. Zhang M, Barash S. Neuronal switching of sensorimotor transformations for antisaccades. Nature. 2000; 408:971–5. [PubMed: 11140683]
23. Gail A, Andersen R. a Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations. J Neurosci. 2006; 26:9376–84. [PubMed: 16971521]
24. 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; 110:2653–8. [PubMed: 23345447]
25. Oller DK, Eilers RE, Oiler DK. The Role of Audition in Infant Babbling The Role of Audition in Infant Babbling. Child Development. 1988; 59:441–449. [PubMed: 3359864]
26. Agnew ZK, McGettigan C, Banks B, Scott SK. Articulatory movements modulate auditory responses to speech. NeuroImage. 2013; 73:191–9. [PubMed: 22982103]

27. Goodglass, H.; Kaplan, E.; Barresi, B. Assessment of aphasia and related disorders. Lippincott Williams & Wilkins; 2000.
28. Damasio H, Damasio AR. The anatomical basis of conduction aphasia. Brain. 1980; 103:337–50. [PubMed: 7397481]
29. Benson DF, Sheremata WA, Bouchard R, Segarra JM, Price D, Geschwind N. Conduction aphasia: A clinicopathic study. Arch Neurol. 1973; 28:339–346. [PubMed: 4696016]
30. Murphy K, et al. Cerebral areas associated with motor control of speech in humans Cerebral areas associated with motor control of speech in humans. J Appl Physiol. 1997; 83:1438–1447. [PubMed: 9375303]
31. Yang AI, et al. Localization of dense intracranial electrode arrays using magnetic resonance imaging. NeuroImage. 2012; 63:157–165. [PubMed: 22759995]
32. Kovalev D, et al. Rapid and fully automated visualization of subdural electrodes in the presurgical evaluation of epilepsy patients. AJNR Am J Neuroradiol. 2005; 26:1078–83. [PubMed: 15891163]
33. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I: Segmentation and Surface Reconstruction. NeuroImage. 1999; 9:179–194. [PubMed: 9931268]
34. Mitra P, Pesaran B. Analysis of Dynamic Brain Imaging Data. Biophys J. 1999; 76:691–708. [PubMed: 9929474]
35. Maris E, Schoffelen JM, Fries P. Nonparametric statistical testing of coherence differences. J Neurosci Methods. 2007; 163:161–75. [PubMed: 17395267]
36. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc Ser B (Methodological). 1995; 57:289–300.

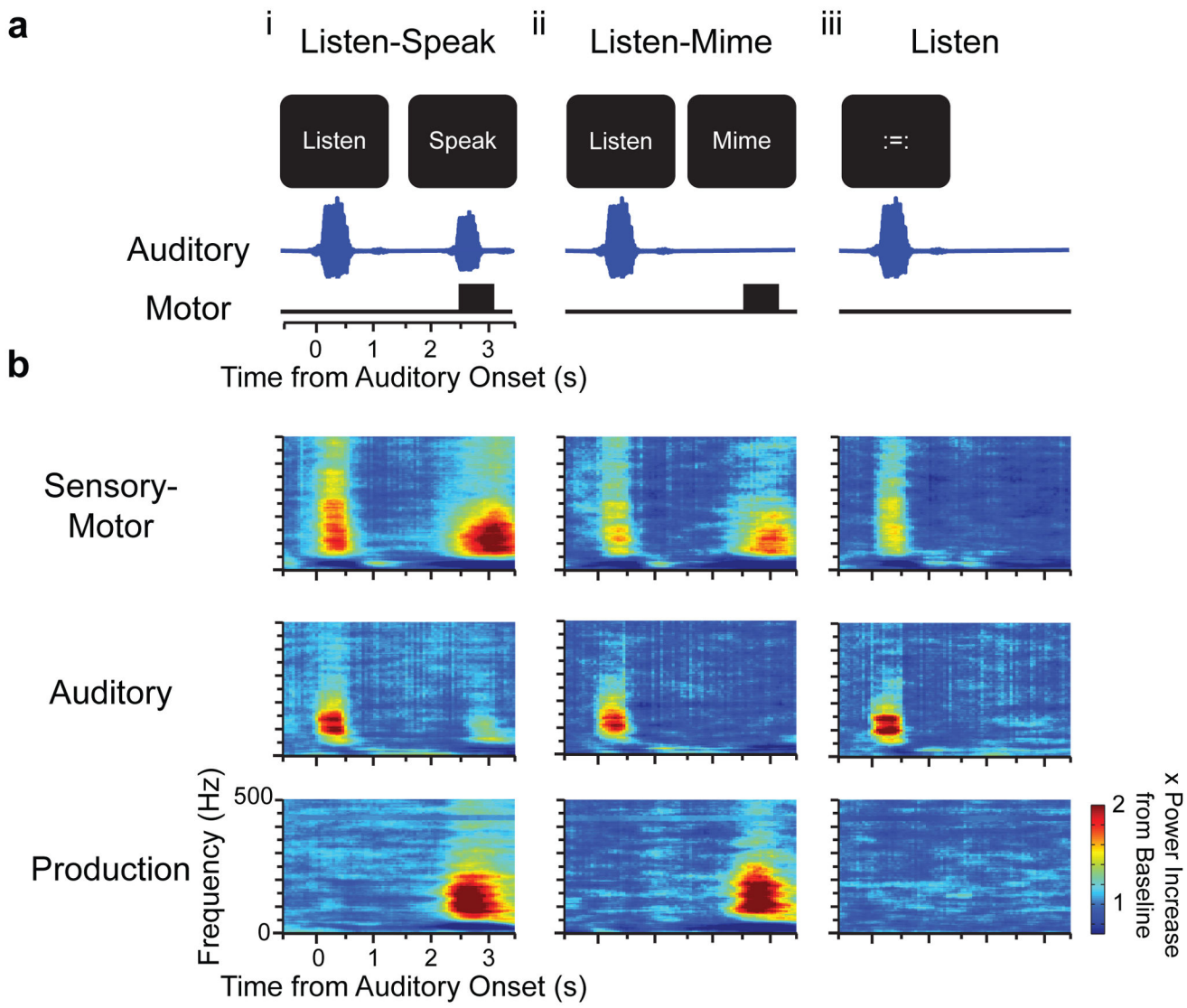

Figure 1. Behavioral tasks and example neural activations

a) Subjects were auditorily presented with a CVC single syllable word and instructed to perform one of three tasks on interleaved trials: **i.** Listen-Speak - Listen to the word, visual prompt 'Listen'. After a 2 s delay repeat the word, visual prompt 'Speak'. **ii.** Listen-Mime – Listen as for **ii**. After delay, mime speaking the word, visual prompt 'Mime'. **iii.** Listen - Passively listen to the word, visual prompt ':=:'. Auditory and motor timelines (below). **b)** Example time-frequency spectrograms of ECoG activity normalized at each frequency to the Baseline power during visual prompt. **Sensory-motor (S-M):** Significant activity during the auditory and movement epochs in Listen-Speak and Listen-Mime tasks. **Production (PROD):** Significant activity during both movement epochs. **Auditory (AUD):** Significant activity during each task epoch with auditory stimuli.

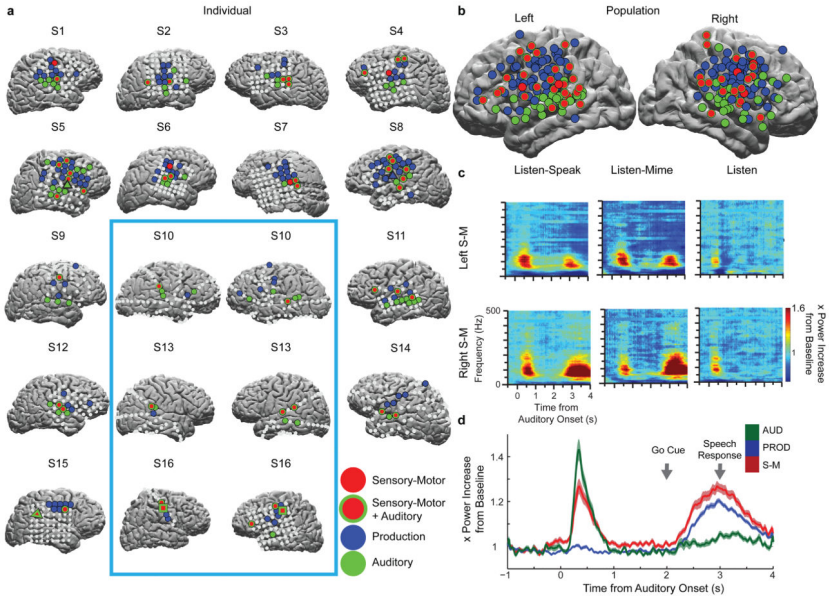

Figure 2. Topography of neural responses and bilateral activation

a) Significant task-related activations within individual subject brains for left (S3,S4,S7,S8,S11,S14), right (S1,S2,S5,S6,S9,S12,S15), or both (S10,S13,S16) hemispheres. Bilateral coverage (light blue box). Electrodes with significant high gamma activity (70–90 Hz): AUD (green), PROD (blue), and S-M (red). AUD and S-M activations (red with green) were often present on the same electrode. Electrodes without significant activation (grey). Triangles denote example activations from Fig 1b, and squares (S16) denote example spectrograms in Fig 2c. b). Significant electrodes projected onto population average left and right hemispheres, color convention as **a)**. Electrode sizes have been increased for illustrative purposes (actual sizes - Fig S4. **c)** Neural spectrograms for example S-M electrodes in left (upper) and right (lower) hemispheres of S16 during Listen-Speak, Listen-Mime, and Listen tasks. **d).** Population average neural response profiles for each class of electrodes. Shaded regions indicate SEM values across electrodes. Go Cue and average production response onset (grey arrows).

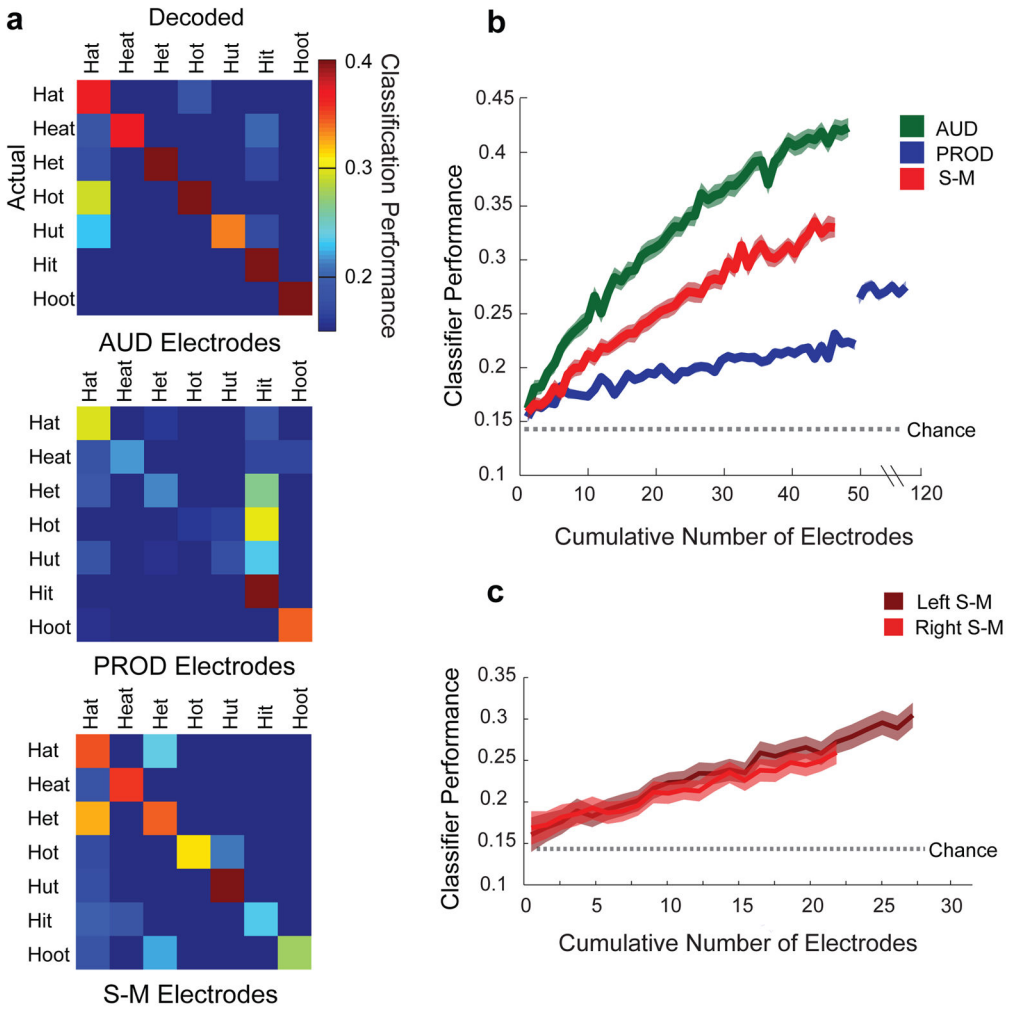

Figure 3. Neural decoding of words

a) Confusion matrices for a 7-way linear classifier using neural responses. AUD electrodes (top panel). PROD electrodes (middle panel). S-M electrodes (bottom panel). Performance is thresholded at chance performance, p = 0.14, for display purposes only. **b)** Classification performance for increasing numbers of electrodes. Chance performance (dotted). **c)** Classification performance for S-M electrodes in the left (dark red) and right (light red) hemispheres. Chance performance (dotted). **Online Methods** presents S-M results by response epoch.

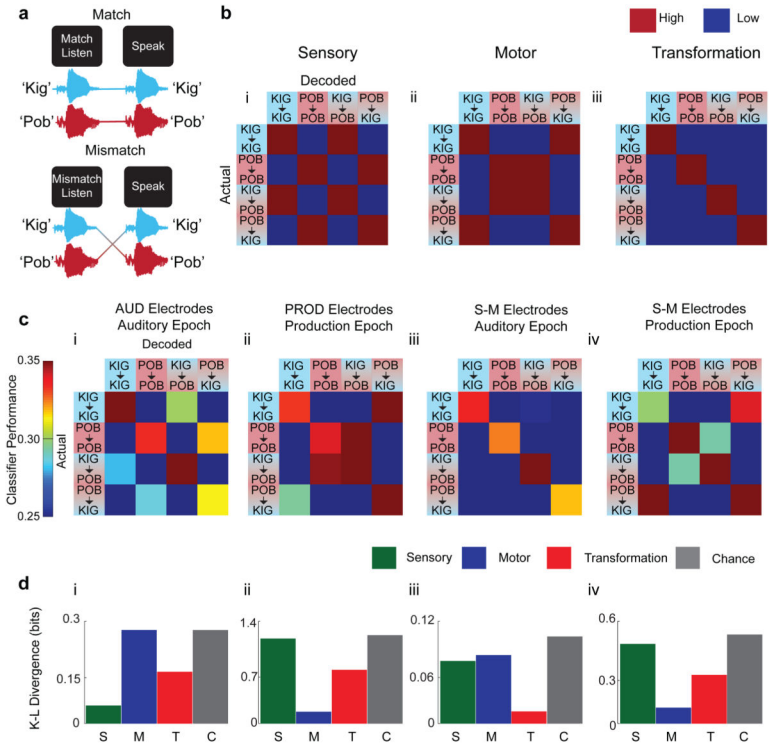

Figure 4. Listen-Speak Transformation task

a) In the Listen-Speak Transformation task, subjects have to transform a non-word they hear into a non-word they speak according to a simple rule. Subjects were first presented with a visual cue: 'match listen' or 'mismatch listen' that instructed the rule that determined the non-word to say in response to the word they heard. On 'match', trials the rule was to repeat the non-word they heard. On 'mismatch' trials, they should say the non-word they did not hear. The non-words were 'kig' and 'pob'. Subjects then heard one of the two non-words, waited for a short delay, then said the appropriate non-word in response to the 'Speak' cue. There were four task conditions. Kig→kig: hear 'kig' and say 'kig'. Pob→pob: hear 'pob' and say 'pob'. Kig→pob: hear 'kig' and say 'pob'; and Pob→kig: hear 'pob' and say 'kig'. **b)** Confusion matrices predicted by the Sensory, Motor and Transformation models. **c)** Confusion matrices during the Listen-Speak Transformation task. **d)** Model fit quantified using a Kullback-Leibler (K-L) divergence.