Published in final edited form as:
Prog Brain Res. 2013 ; 207: 435–456. doi:10.1016/B978-0-444-63327-9.00018-7.

Decoding Speech for Understanding and Treating Aphasia

Brian N. Pasley[*,1] and **Robert T. Knight**[*,†,‡]

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

[†]Department of Neurological Surgery, University of California—San Francisco, San Francisco, CA, USA

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

Abstract

Aphasia is an acquired language disorder with a diverse set of symptoms that can affect virtually any linguistic modality across both the comprehension and production of spoken language. Partial recovery of language function after injury is common but typically incomplete. Rehabilitation strategies focus on behavioral training to induce plasticity in underlying neural circuits to maximize linguistic recovery. Understanding the different neural circuits underlying diverse language functions is a key to developing more effective treatment strategies. This chapter discusses a systems identification analytic approach to the study of linguistic neural representation. The focus of this framework is a quantitative, model-based characterization of speech and language neural representations that can be used to decode, or predict, speech representations from measured brain activity. Recent results of this approach are discussed in the context of applications to understanding the neural basis of aphasia symptoms and the potential to optimize plasticity during the rehabilitation process.

Keywords

aphasia; speech; language; neural encoding; decoding

1 INTRODUCTION

Aphasia is a language disorder resulting from brain damage, often to frontotemporal cortex, that causes behavioral deficits in the production or comprehension of speech. Aphasic symptoms affect a diversity of language components, including auditory, phonological, or lexical function. This diversity represents a challenge for development of effective treatments that must target specific brain circuits underlying the heterogeneity of language abilities.

Treatment of aphasia relies on behavioral interventions that encourage structural and/or functional brain plasticity for recovery of language ability. The rehabilitation process may be aided by maximizing neural plasticity in language areas to recover damaged circuits.

[1]Corresponding author: Tel.: +510-643-9744; Fax: +510-642-3192, bpasley@berkeley.edu.

Defining the basic neural mechanisms supporting specific language functions is an important building block for clinical insights that can help to identify injured cortical systems and to improve targeted treatments.

2 APHASIA SUBTYPES, SYMPTOMS, AND REHABILITATION

Aphasia is a linguistic disorder and deficits occur in virtually all modalities. For instance, an aphasic patient with inability to speak would also have impaired writing ability and a patient who cannot comprehend speech would also have deficits in reading capacity. Broca's aphasia is typically due to damage in the left inferior frontal gyrus and nearby subcortical structures including the anterior insula. Classic Broca's aphasia patients have preserved comprehension and varying degrees of inability to produce language that are evident in all modalities of language production. Damage in left posterior temporal lobe including the superior temporal plane, the superior temporal gyrus (STG), and the middle temporal gyrus results in Wernicke's aphasia and problems with speech comprehension. Lesions to the arcuate fasciculus connecting Broca's and Wernicke's regions cause problems in repetition of auditory information with largely intact production and comprehension of speech. However, it should be noted that speech output deficits can be observed in posterior temporal lesions and prominent comprehension deficits can be seen on inferior frontal lesions.

Recovery to some degree is seen in most aphasics. However, patients with extensive recovery of linguistic skills are often left with anomia, which is the inability to name objects despite knowing what they are and how to use them. Extensive damage to the left hemisphere perisylvian language cortices including Broca's and Wernicke's area causes the devastating syndrome of global aphasia where the patient cannot speak or understand language. In essence, the global aphasia patient is completely cut off from linguistic interactions with the world. Age of injury is the most salient factor in recovery of function. For instance, a 6-year-old with a global aphasia due to massive perisylvian damage will show massive language improvement and may even look normal at a year post injury. This recovery is presumed due to the engagement of the right hemisphere and highlights the remarkable plasticity of the younger brain. Conversely, a 60-year-old with the same extent of injury and global aphasia will likely remain severely impaired 10–20 years after the injury. Individual hemispheric organization for language is another critical factor in recovery. Over 95% of males and 90% of females are left hemisphere-dominant for language. The remaining right-handed subjects may show bilateral language capacity and over 30% of left-handed subjects have bilateral language representation. Left-handed subjects are also more likely to have bilateral speech representation. From a clinical perspective, patients who have bilateral language representation will have more rapid recovery from aphasia.

In sum, these observations indicate that the location of brain damage largely determines the type of aphasic symptoms that arise. This is consistent with the view of the brain's language system as a modular network, with specific language functions organized into functionally distinct neural circuits (Friederici, 2011; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). However, the complexity and interdependence of language functions, and that of

aphasic symptoms, also suggests a more distributed system where language abilities are supported by multiple, interconnected brain regions (Friederici, 2011; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009).

To begin to understand the link between brain injury and the variety of aphasic symptoms, and how function can be recovered through rehabilitation and reorganization of the underlying circuits, experimental and analytic tools from systems neuroscience can be brought to bear. This chapter focuses on a systems identification approach that, by use of neural encoding and decoding models, characterizes how neural activity relates to heterogeneous aspects of speech. The advantage of this approach is it offers a quantitative, model-based description of the speech features encoded by specific neural circuits, providing insights into the dependence of aphasic symptoms on the location of injury and potential treatment avenues.

3 A NEURAL SYSTEMS APPROACH TO LANGUAGE

3.1 Functional Organization of Language

The brain's language system can be divided broadly into two areas supporting speech comprehension and speech production. Classically defined as Wernicke's and Broca's areas, as briefly reviewed earlier, damage to these regions can lead to receptive or expressive aphasic symptoms, respectively. Both speech comprehension and production are believed to involve multiple stages of neural representation (Friederici, 2011; Hackett, 2011; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009).

For speech comprehension, the brain's task is to convert "sound to meaning." The first stages in this process involve acoustic signal analysis in early auditory cortex. At the highest level of analysis, the brain computes semantic representations concerned with word meanings. A hierarchy of cortical areas underlies this complex transformation, which maps incoming low-level acoustic sounds to intermediate, categorical representations and ultimately to high-level neural representations of semantic meaning (Friederici, 2011; Hackett, 2011; Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). Anatomically, the so-called auditory "what" pathway has been reported to extend along an anterolateral gradient in superior temporal cortex (Rauschecker and Scott, 2009) where stimulus selectivity increases from pure tones in primary auditory cortex to words and sentences in anterior temporal cortex (Friederici, 2011).

For speech production, articulatory representations are likely coded in the frontal lobe within motor, premotor, and Broca's areas (Bouchard et al., 2013; Goense and Logothetis, 2008; Hickok and Poeppel, 2007; Sahin et al., 2009; Tankus et al., 2013). At the lowest level, these areas may code the activation of individual muscles within the vocal tract (Lofqvist, 1999) that control the complex sequence of movements during articulation. At a higher level, speech motor control may involve coding of entire "gestures" or coordinated muscle synergies (Graziano and Aflalo, 2007; Lofqvist, 1999). Higher-level linguistic functions such as lexical, grammatical, and phonological information may also be coded in Broca's area (Sahin et al., 2009).

Within this broad framework, the neural representation of language remains a difficult experimental question that has resisted precise delineation for over 150 years. Animal models have been widely explored in nonhuman mammals and avians in the context of lower-level auditory and motor processing (Aertsen and Johannesma, 1981; De Boer, 1967; deCharms et al., 1998; Depireux et al., 2001; Georgopoulos et al., 1982; Theunissen et al., 2001; Todorov, 2004). However, extending these findings to higher-level speech processing in humans has been impeded because fine-scale, invasive recording opportunities are limited. In general, this experimental barrier, and the intricacy of human language ability, has made it difficult to develop effective neural models of language that approach the detail and predictive accuracy achieved by existing animal models. Speech is unique to humans, and there are significant specializations that have evolved for speech processing in the human brain that cannot be readily studied in animal models.

3.2 Electrocorticography (ECoG)

Recordings in the human brain are generally restricted to noninvasive techniques such as electroencephalography (EEG), magnetoencephalography (MEG), or functional magnetic resonance imaging (fMRI). These techniques have yielded key insights into large-scale language organization (Formisano et al., 2008; Mitchell et al., 2008; Schonwiesner and Zatorre, 2009) but have less spatial and temporal resolution than traditional invasive microelectrode recordings available in animal models. In rare cases, direct electrode recordings can also be obtained in human patients who are undergoing neurosurgical procedures for epilepsy or brain tumor. In such cases, clinical treatment requires temporary implantation of subdural electrode arrays onto the cortical surface. These intracranial electrocorticographic (ECoG) recordings represent a unique opportunity to obtain neural recordings from broad areas of language-related cortex at high spatiotemporal resolution (millimeter and millisecond scale). The ECoG signal measured by individual electrodes is a neural signal similar to the cortical local field potential (LFP). In particular, this chapter focuses on the high-gamma component (70–150 Hz), a population-level signal, which has been shown, in LFP recordings to correlate with multiunit spike rate of the local neuronal population (Goense and Logothetis, 2008; Viswanathan and Freeman, 2007). Previous intracranial studies (Canolty et al., 2007; Crone et al., 2001; Edwards et al., 2009; Nourski et al., 2009; Pei et al., 2011a,b) have found that speech perception and production evoke robust and sustained increases in high-gamma band power in temporal, frontal, and parietal cortices. This chapter focuses on recent work from intracranial human recordings that seeks to bridge the gap between detailed animal models of low-level auditory-motor processing and relatively unexplored models of intermediate- and higher-level speech processing in humans.

3.3 Neural Encoding and Decoding Models

To study cortical speech representation, an effective analytic approach is the use of neural encoding predictive models. A neural encoding model characterizes the relationship between speech function and measured brain activity. It describes the stimulus or behavioral features that account for and accurately predict the neural response. For example, does neural activity in STG encode acoustic parameters of speech or does it code entire categories like consonants and vowels? Does Broca’s area encode muscle movements or higher-level

syntactic rules of language? The use of encoding models to test such hypotheses offers a quantitative answer to how well observed neural responses are described by each hypothesis (Figs. 1–4).

In this analytic framework, data are assumed to be generated by a black box that takes as input a set of predictor (independent) variables and outputs a set of response (dependent) variables. The black box represents nature's true relationship between the predictor and response variables, and the goal is to develop statistical models that emulate nature's system as closely as possible (Breiman, 2001). Hypothesis testing is implicit in the ability of the statistical model to predict new data (i.e., emulate nature). Different encoding models encapsulate different hypotheses about speech function. These hypotheses are tested by comparing the predictive power of the encoding models, with a "perfect" model yielding perfect predictions. Because prediction is rarely perfect, the model residuals can be examined to identify particular aspects of the data not accounted for by the model. This in turn provides specific information to formulate new models and offers a principled and structured approach to iterative hypothesis testing.

Encoding models characterize the forward transformation from stimulus to response. In its basic form, this is a many-to-one transformation that maps multiple inputs (e.g., stimulus features) onto a single output (e.g., the neural response from one electrode/sensor). The resulting statistical model describes estimates of neural tuning—the responsivity of the neural response to different stimulus features. For example, the spectrotemporal receptive field (STRF) model, ubiquitous in the study of early auditory cortex (Aertsen and Johannesma, 1981; De Boer, 1967; deCharms et al., 1998; Depireux et al., 2001; Theunissen et al., 2001), describes the frequency selectivity of individual neurons, that is, the observation that the responses of single auditory neurons prefer a narrow range of sound frequencies peaking at low, middle, or high values in the acoustic spectrum (Figs. 2 and 3).

It is often useful to study stimulus representation encoded by an entire neuronal population, as measured from multiple sensors rather than a single neural response. A related approach to studying neuronal population responses is the reverse transformation, that is, a "decoding" model that maps neural responses to stimulus. This is a many-to-many mapping that uses the measured neural responses to predict or decode the stimulus features under study. For example, in brain–machine interface applications, decoding models have been used to predict the position of a computer cursor by learning the mapping from a population of motor neurons to the direction and velocity of the cursor on the screen (Carmena, 2013). Quantifying which stimulus or behavioral features are successfully decoded or reconstructed from population activity reveals which aspects of the stimulus are encoded by the neuronal population as a whole.

Neural encoding and decoding models are central to sensory neurophysiology (Wu et al., 2006) and brain–machine interface (Carmena, 2013). Recent work has also demonstrated how this approach can be usefully applied to study different aspects of speech or language in the human cortex (Brumberg et al., 2010; Tankus et al., 2013). Multiple levels of speech representation have been successfully decoded using intracranial neural signals. These include auditory representations (Guenther et al., 2009; Pasley et al., 2012), consonants and

vowels (Pei et al., 2011a,b; Tankus et al., 2012), and words (Kellis et al., 2010). Later, we will review a number of these results as applied to three different levels of speech representation: auditory, phonetic, and articulatory processing.

3.4 Spectrotemporal Encoding in Auditory Cortex

In recent work, we investigated the neural representation of speech by measuring ECoG responses to natural speech in clinical patients undergoing treatment for epilepsy (Pasley et al., 2012). We tested the ability of two different auditory encoding models to explain measured ECoG responses from the STG, a nonprimary auditory area. These models are based on decades of research on the response properties of neurons in the mammalian auditory system. The emerging frame-work of early auditory processing consists of two conceptually similar stimulus transformations (Chi et al., 2005; Dau et al., 1997; Depireux et al., 2001; Eggermont, 2002; Miller et al., 2002) (Fig. 3). In the first, an auditory filter bank extracts spectral energy from the one-dimensional sound pressure waveform, essentially building an auditory spectrogram representation of the sound. The spectrogram model is based on the spectrotemporal envelope of the speech stimulus. This model assumes that neural responses are a linear function of spectrotemporal auditory features and are equivalent to the standard STRF (Aertsen and Johannesma, 1981; De Boer, 1967; deCharms et al., 1998; Depireux et al., 2001; Theunissen et al., 2001) In the second stage, a modulation-selective filter bank analyzes the two-dimensional auditory spectrogram and extracts modulation energy at different temporal rates and spectral scales (Chi et al., 2005). The key advantage of this representation, referred to here as the “modulation model,” is that it *explicitly* represents amplitude envelope modulations that have a fundamental relationship with the information-bearing components of speech. This representation emphasizes robust features of speech that correspond to, for example, formants in the spectral axis, syllable rate in the temporal axis, and formant transitions in the joint spectrotemporal space (Chi et al., 2005). In contrast to fine spectrotemporal acoustic structure, which may exhibit significant variability under natural conditions, these slow, relatively coarse patterns in modulation space carry essential phonological information and are correlated with psychophysical measures of speech intelligibility and are robust under a variety of noise conditions (Chi et al., 1999; Dau et al., 1997; Elliott and Theunissen, 2009).

To investigate the neural representation of speech in STG, we compared the predictive power of forward encoding models that predicted high-gamma activity from the auditory stimulus based on either spectrogram or modulation representation (Fig. 4). Model parameters are fitted directly to cortical responses to natural speech and predict neural activity to novel stimuli not used in the fitting procedure. The fitted parameters describe neural tuning to acoustic frequency, spectral modulation (scale), and temporal modulation (rate).

Across responsive electrodes, predictive power for the modulation model is slightly better compared to the spectrogram model (Fig. 4). The fitted spectrogram-based models exhibit a complex tuning pattern with multiple frequency peaks (Figs. 1 and 4). Linear STRFs of peripheral auditory single neurons exhibit only a single “best frequency” (Schreiner et al., 2000), although multipeaked frequency tuning has been observed in several auditory areas

(Kadia and Wang, 2003; Rauschecker et al., 1997; Sutter and Schreiner, 1991). Multipeaked tuning at surface electrodes may therefore reflect spatial integration of a range of individual best frequencies in the underlying neurons or, alternatively, large neuronal populations with higher-order selectivity for frequency conjunctions.

To quantify how the population of neuronal responses across STG encodes the spectrogram and modulation speech representations, we also used decoding models to reconstruct these representations from multielectrode ECoG responses (Bialek et al., 1991;Mesgarani et al., 2009). We decoded the original stimulus from cortical activity patterns at multiple electrodes using a regularized linear regression algorithm similar to previous motor brain–machine interface (Carmena et al., 2003) or sensory experiments (Bialek et al., 1991; Mesgarani et al., 2009). This method places an upper bound on coding accuracy by quantifying the fidelity with which specific features are encoded in the cortical population response. Reconstructed stimuli were compared directly to the original speech representation (Fig. 5). We found that both the spectrogram-based and modulation-based representations can be accurately decoded from single-trial brain activity. However, a key difference is that the modulation-based reconstruction exhibits substantially higher fidelity for rapid modulations of the amplitude envelope, while fidelity of the spectrogram-based reconstruction has a low-pass characteristic (Fig. 5). Rapid spectral modulations comprising vowel harmonics are clearly isolated by the modulation-based reconstruction but are not fully resolved in the spectrogram-based reconstruction (Fig. 5). Similarly, the modulation-based reconstruction more clearly resolves rapid temporal modulations such as those distinguishing syllable onsets and offsets (Fig. 5). The distinction between spectrogram and modulation reconstruction is evident when accuracy is assessed component-by-component across all subjects (Fig. 5). The modulation-based model recovers the full space of spectrotemporal modulations, while the spectrogram-based model fails to decode higher temporal rates (> 4 Hz) and spectral scales (> 2 cyc/oct at rates >2 Hz) (Fig. 5). The enhanced reconstruction quality in the modulation energy domain suggests that the energy representation provides a better functional description of the stimulus–response transformation in higher-order auditory cortex.

3.5 Phonetic Encoding in Auditory Cortex

During natural speech, communication requires the extraction of meaning from a highly variable acoustic signal. Variability in speech sounds arises from many sources, including differences among speakers (male or female pitch), tempo (slow vs. fast speaking rates), and dialect (Greenberg, 2006; Greenberg and Ainsworth, 2004). Further variability is introduced by contextual effects (coarticulation) between adjacent phonemes, the basic units of speech that convey meaning. Consonant and vowel phonemes may have unique phonetic identities, but individual acoustic realizations originate from a distribution of speech sounds, the spectral properties of which may bear little similarity across examples. For instance, in Fig. 6, the vowel [ux] is spoken twice during the sentence. Acoustically, these two utterances differ in spectrotemporal content, exhibiting different formant dynamics as a consequence of the adjacent phones (coarticulation). Utterances from two different speakers would exhibit even greater spectral differences. Phonetically, however, these two utterances are considered to represent the identical category (i.e., the vowel [ux]) irrespective of acoustic differences.

This central problem of acoustic variance has led to the suggestion that a precisely accurate representation of the acoustic signal might actually impede intelligibility, particularly in the face of external factors such as background noise or competing speech (Greenberg, 2006; Greenberg and Ainsworth, 2004). Given the ease by which humans communicate across diverse and challenging environmental conditions, the human auditory system solves the problem of acoustic variability with remarkable efficiency. How does the human brain build invariant categorical representations of highly variable acoustic signals in order to extract meaning? How does brain injury alter this process and lead to language impairments observed in aphasia?

To investigate phonetic neural representation, we examined the pattern of neural responses to continuous streams of natural speech, which contain characteristic sequences of consonant–vowel (CV) patterns (Greenberg, 2006). In an auditory stream of speech, words and syllables function as discrete units, yet the sound itself is continuous. Speech comprehension depends on segmentation cues that allow listeners to segment continuous speech sounds into meaningful phonetic units, for instance CV syllables. What do the auditory encoding models described in the previous section reveal about the neural basis of this segmentation process? To address this question, we investigated the relationship between patterns of stimulus tuning in the encoding models and the average cortical response to consonants and vowels embedded in phonetically transcribed English sentences (Garofolo et al., 1993). First, we found a number of electrode sites that showed a robust high-gamma response to vowels compared to consonants (Fig. 7). The distinct spectrotemporal properties of consonants and vowels (Mesgarani and Shamma, 2011; Mesgarani et al., 2008) suggested a possible basis for this response selectivity. In particular, consonants are transient sounds with rapid onset/offset, activating high temporal rates (>8 Hz). In contrast, vowels are characterized by a fast onset of harmonic structure (activating high rates) that persists at relative steady state for the duration of the vowel, activating intermediate rates (2–8 Hz) (Mesgarani and Shamma, 2011; Mesgarani et al., 2008). Notably, modulation tuning observed across the full electrode ensemble was well matched to these intermediate rates (~2–8 Hz, Fig. 7). This suggests that modulation tuning might explain the observed sensitivity to consonants versus vowels. To further examine if neural tuning patterns can account quantitatively for sensitivity to vowels versus consonants, we used the estimated models to filter a large set of natural speech stimuli and assessed the average predicted CV response selectivity (David et al., 2006). For each site, we compared the measured CV response selectivity to that predicted by the fitted models. The measured high-gamma CV response difference is strongly correlated with that predicted by estimated modulation models (Fig. 7, r=0.77, $p < 10^{-7}$). Selectivity for vowel like sounds can therefore be explained in part by the modulation tuning measured at specific cortical sites. This finding suggests an interesting possibility that vowel-sensitive sites in higher-order auditory cortex may participate in the process of syllable segmentation by detecting the presence of vowel like structures, which, in many languages, comprise the syllable nucleus (Greenberg, 2006).

Recent work has also demonstrated that categorical information about CV syllables can be decoded directly from STG (Chang et al., 2010). Using a classic psychophysical paradigm

(Liberman et al., 1967), Chang and Rieger et al. (Chang et al., 2010) measured ECoG activity in the STG during auditory presentation of three CV syllables, /ba/, /da/, and /ga/. In this paradigm, a series of stimuli are synthesized that vary continuously in the starting frequency of the F2 transition (second vocal tract resonance) such that the listener's perception varies across three initial consonants from /ba/, to /da/, to /ga/. Although the actual acoustic parameter, the starting F2 frequency, varies continuously, the perception of the listener is discrete, corresponding to one of the three discrete CV categories. Using this dissociation between perception and physical stimulus, Chang and Rieger (Chang et al., 2010) identified neural signals that encoded the categorical perception, as opposed to the continuous acoustic parameter. A classification algorithm was used to decode CV category from ECoG signals recorded across STG. The results revealed a distributed representation of STG sites that allowed the classifier to accurately decode the subject's categorical percept as opposed to the continuous physical stimulus (Fig. 8). The findings reveal that the use of encoding and decoding models, as described in these examples, can help identify important stimulus features coded by the neural system and the distribution of cortical sites that support the given function or behavior.

3.6 Articulatory Encoding in Motor Cortex

Encoding and decoding models have also been applied to the neural basis of speech production. Speech motor control involves steps to select specific articulator muscles, establish the degree of activation in each muscle, and initiate a coordinated activation sequence. A central question in motor control is whether neurons in primary motor cortex represent low-level parameters of movement such as muscle activations or, alternatively, high-level aspects such as movement goals (Graziano and Aflalo, 2007; Todorov, 2004). This open question, "muscles or movements?," is important for investigating the role of motor cortex in speech production. For example, during the articulation of individual phonemes, a low-level representation would predict topographic cortical activation corresponding to the engaged articulators. A high-level representation might predict more general building blocks that correspond to groups of muscles or, at an extreme, distinct neural circuits devoted to articulation of each phoneme.

For speech production, one possible encoding model is based on the pattern of muscle activation in a set of speech articulators across time. In this example, we focus on the major articulators that have robust cortical representations in the motor homunculus, including the lips, tongue, and larynx. The basic premise of this articulator-based representation is the observation that different phonemes have distinct temporal patterns of coordinated articulator movement (Lofqvist, 1999). The specific temporal sequence for a given utterance is commonly referred to as a "gestural score" (Browman and Goldstein, 1989). Across time, individual articulators become active and inactive in a coordinated fashion to produce specific phonetic signals. Articulatory phonology makes the key assumption that phonological contrast in speech can be defined in terms of different gestural scores. This assumption is supported by articulatory measurements where, to a large extent, there is a one-to-one mapping from the physical configuration of the vocal tract's articulators to the phoneme (Deng and O'Shaughnessy, 2003). An articulatory-based encoding model takes advantage of this direct correspondence. Specifically, the model uses one input feature for

each individual articulator, with a simple on/off coding for whether or not the articulator is active at each time point. This model assumes that motor neural activity is a linear function of temporal patterns of muscle activation in a set of articulators. In a sense, the neural activity serves as a first-order proxy for the underlying gestural score, which can then be used to decode individual phonemes.

To determine average gestural scores for individual phonemes, we used the MOCHA-TIMIT speech corpus (Wrench and Hardcastle, 2000), which includes simultaneous acoustic and articulatory measurements obtained from electromagnetic articulography (Fig. 9). The time-stamped phonetic transcription can be used to derive linear estimates of the various articulator activities during the articulation of each phone. Figure 9 shows the articulatory impulse response of three different phones. As expected, the bilabial consonants /b/ and /p/ have similar responses focused on the upper and lower lips. In contrast, the dental consonant /l/, which is articulated with a flat tongue against the alveolar ridge and upper teeth, exhibits a strong response in the tongue foci.

We next investigated the relationship between motor cortex ECoG activity and the average gestural scores of speech articulators during phoneme production. Figure 9 shows, for a single patient, the cortical motor representation of three major articulators, the lips, tongue, and larynx. These sites are determined by electrical cortical stimulation mapping in which stimulation is applied to evoke movements in order to map out the motor cortex for presurgical evaluation. Figure 9 shows that individual articulator dynamics are represented in motor cortex ECoG activity as patients read aloud visually displayed monosyllables, such as /ba/, /pa/, and /la/. Time zero indicates the acoustic onset of the spoken syllable as determined from the audio recording. For /ba/ and /pa/, high-gamma activity in the lip electrode (blue curve) increases prior to acoustic onset, while tongue activity (red curve) is flat. This is qualitatively consistent with the gestural score for bilabial consonants. Similarly, for /la/, activity in the tongue electrode (red) increases prior to acoustic onset, while lip activity remains relatively flat. Interestingly, activity in the larynx electrode (green curve) is robust and remains elevated for the duration of all three syllables. This is likely due to the abduction and adduction of the laryngeal muscles during preparation and maintenance of voicing onset (Hajime, 1999).

To directly evaluate this simple encoding model, the muscle activity of individual articulators would need to be measured simultaneously with ECoG neural signals, a difficult experimental setup. Nevertheless, the qualitative comparison offered here illustrates the general usefulness of an encoding model approach to the neural basis of speech production. In principle, the predictive power of this first-order articulator model could be compared directly to alternative encoding models that propose higher-level movement representations incorporating second- and higher-order interactions between articulators. For example, recent evidence from sensorimotor cortex suggests the existence of a phonetically organized gesture representation during speech articulation (Bouchard et al., 2013).

4 APPLICATIONS TO APHASIA

A systems identification approach to investigate different representation levels in the language system provides a principled experimental framework for study of underlying neural mechanisms. Encoding and decoding models can be used to identify speech features or language rules that are represented by distributed neural activity in language-related cortex. Prediction accuracy of alternative models can be compared to test hypotheses about which representations best explain measured neural activity.

A better understanding of the neural mechanisms underlying different language functions will help identify injured neural circuits in aphasic patients and potentially suggest targeted strategies to induce neural plasticity during rehabilitation. For example, recent work (Robson et al., 2013) identified impairments in basic spectrotemporal modulation processing of auditory stimuli in Wernicke's aphasia patients with lesions to parietal and superior temporal areas. Notably, these same areas, STG and parietal cortex, have well-defined patterns of modulation tuning, as described earlier (Fig. 4) and in Pasley et al. (2012). It is possible that disruption to this modulation tuning underlies the observed auditory impairments. In this case, aphasic symptoms appear to have a close relationship with the underlying speech representation in the lesioned cortical areas.

While this example offers a possible functional explanation for specific aphasic symptoms, a more powerful application of the systems identification approach would be to directly measure changes in neural representation induced by rehabilitation. For example, neural encoding models could be estimated continuously during the rehabilitation process to characterize changes in modulation tuning induced by treatment procedures. Lack of change in the underlying neural tuning would indicate an ineffective treatment. On the other hand, observed increases to modulation sensitivity could be used to optimize and guide training-induced plasticity. With accurate encoding models for each level of cortical speech representation, the same approach would be applicable to a variety of aphasic symptoms, such as impairments to phonetic, semantic, or articulatory processing.

Although potential clinical insights into aphasia are evident, the use of encoding and decoding models as a diagnostic or treatment tool has many important experimental challenges that are currently unmet. For example, detailed models of speech have been estimated primarily using invasive recording methods that are in general neither available nor appropriate for aphasia patients. Monitoring plasticity in underlying neural speech representations using this tool is therefore not currently feasible. However, recent work demonstrates that detailed encoding and decoding models are possible with other noninvasive methods including fMRI (Naselaris et al., 2009; Nishimoto et al., 2011). In the visual system, such models have been extensively applied to characterize the neural representation of natural images in numerous visual areas and to decode the visual content of dynamic visual movies (Nishimoto et al., 2011). This work demonstrates that fMRI has sufficient spatial resolution to provide detailed characterizations of neural tuning. In the auditory system, fMRI has been used to detect patterns of frequency and modulation tuning by characterizing tonotopic maps (Talavage et al., 2004) and modulation sensitivity in different auditory areas (Schonwiesner and Zatorre, 2009). Models derived from fMRI have

also provided insights into the larger-scale distributed representation of phonemes (Formisano et al., 2008) and semantic properties of nouns (Mitchell et al., 2008). Important challenges remain, for example, how to use fMRI, which has a temporal resolution on the order of seconds, to capture the rapid temporal dynamics of speech (on the order of milliseconds). Despite these challenges, ongoing research to improve temporal resolution in fMRI or to combine it with higher-resolution methods such as EEG or MEG offers promising avenues for noninvasive application of neural encoding and decoding models. As these methods improve, opportunities to directly measure plasticity in aphasia rehabilitation may offer novel insights into effective treatment strategies.

References

Adelson EH, Bergen JR. Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A. 1985; 2:284–299. [PubMed: 3973762]

Aertsen AM, Johannesma PI. The spectro-temporal receptive field. A functional characteristic of auditory neurons. Biol. Cybern. 1981; 42:133–143. [PubMed: 7326288]

Bialek W, Rieke F, De Ruyter Van Steveninck RR, Warland D. Reading a neural code. Science. 1991; 252:1854–1857. [PubMed: 2063199]

Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. Nature. 2013; 495(7441):327–332. http://dx.doi.org/10.1038/nature11911. Epub 2013 Feb 20. [PubMed: 23426266]

Breiman L. Statistical Modeling: The Two Cultures. Stat. Sci. 2001; 16:199–231.

Browman CP, Goldstein L. Articulatory gestures as phonological units. Phonology. 1989; 6:201–251.

Brumberg JS, Nieto-Castanon A, Kennedy PR, Guenther F. Brain-computer interfaces for speech communication. Speech Commun. 2010; 52:367–379. [PubMed: 20204164]

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

Carmena JM. Advances in neuroprosthetic learning and control. PLoS Biol. 2013; 11:e1001561. [PubMed: 23700383]

Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, Dimitrov DF, Patil PG, Henriquez CS, Nicolelis MA. Learning to control a brain-machine interface for reaching and grasping by primates. PLoS Biol. 2003; 1:E42. [PubMed: 14624244]

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT. Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. 2010; 13:1428–1432. [PubMed: 20890293]

Chi T, Gao Y, Guyton MC, Ru P, Shamma S. Spectro-temporal modulation transfer functions and speech intelligibility. J. Acoust. Soc. Am. 1999; 106:2719–2732. [PubMed: 10573888]

Chi T, Ru P, Shamma SA. Multiresolution spectrotemporal analysis of complex sounds. J. Acoust. Soc. Am. 2005; 118:887–906. [PubMed: 16158645]

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

Dau T, Kollmeier B, Kohlrausch A. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 1997; 102:2892–2905. [PubMed: 9373976]

David SV, Hayden BY, Gallant JL. Spectral receptive field properties explain shape selectivity in area V4. J. Neurophysiol. 2006; 96:3492–3505. [PubMed: 16987926]

David SV, Mesgarani N, Shamma SA. Estimating sparse spectro-temporal receptive fields with natural stimuli. Network. 2007; 18:191–212. [PubMed: 17852750]

De Boer E. Correlation studies applied to the frequency resolution of the cochlea. J. Audit. Res. 1967; 7:209–217.

Decharms RC, Blake DT, Merzenich MM. Optimizing sound features for cortical neurons. Science. 1998; 280:1439–1443. [PubMed: 9603734]

Deng, L.; O'Shaughnessy, D. Speech Processing: A Dynamic and Optimization-Oriented Approach. New York: Marcel Dekker, Inc.; 2003.

Depireux DA, Simon JZ, Klein DJ, Shamma SA. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J. Neurophysiol. 2001; 85:1220–1234. [PubMed: 11247991]

Edwards E, Soltani M, Kim W, Dalal SS, Nagarajan SS, Berger MS, Knight RT. Comparison of time-frequency responses and the event-related potential to auditory speech stimuli in human cortex. J. Neurophysiol. 2009; 102:377–386. [PubMed: 19439673]

Eggermont JJ. Temporal modulation transfer functions in cat primary auditory cortex: separating stimulus effects from neural mechanisms. J. Neurophysiol. 2002; 87:305–321. [PubMed: 11784752]

Elliott TM, Theunissen FE. The modulation transfer function for speech intelligibility. PLoS Comput. Biol. 2009; 5:e1000302. [PubMed: 19266016]

Formisano E, De Martino F, Bonte M, Goebel R. "Who" is saying "what"? Brain-based decoding of human voice and speech. Science. 2008; 322:970–973. [PubMed: 18988858]

Friederici AD. The brain basis of language processing: from structure to function. Physiol. Rev. 2011; 91:1357–1392. [PubMed: 22013214]

Garofolo JS, Lamel LF, Fisher WM, Fiscus JG, Pallet DS, Dahlgrena NL, Zue V. Acoustic-Phonetic Continuous Speech Corpus. Linguistic Data Consortium. 1993 http://catalog.ldc.upenn.edu/LDC93S1.

Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. J. Neurosci. 1982; 2:1527–1537. [PubMed: 7143039]

Goense JB, Logothetis NK. Neurophysiology of the BOLD fMRI signal in a wake monkeys. Curr. Biol. 2008; 18:631–640. [PubMed: 18439825]

Graziano MS, Aflalo TN. Rethinking cortical organization: moving away from discrete areas arranged in hierarchies. Neuroscientist. 2007; 13:138–147. [PubMed: 17404374]

Greenberg, S. A multi-tier theoretical framework for understanding spoken language. In: Greenberg, S.; Ainsworth, WA., editors. Listening to Speech: An Auditory Perspective. Mahwah, NJ: Lawrence Erlbaum Associates; 2006.

Greenberg, S.; Ainsworth, WA. Speech processing in the auditory system: an overview. In: Greenberg, S.; Ainsworth, WA.; Popper, AN.; Fay, RR., editors. Speech Processing in the Auditory System. New York: Springer-Verlag; 2004.

Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville JA, Panko M, Law R, Siebert SA, Bartels JL, Andreasen DS, Ehirim P, Mao H, Kennedy PR. A wireless brain-machine interface for real-time speech synthesis. PLoS One. 2009; 4:e8218. [PubMed: 20011034]

Hackett TA. Information flow in the auditory cortical network. Hear. Res. 2011; 271:133–146. [PubMed: 20116421]

Hajime, H. Investigating the physiology of laryngeal structures. In: Hardcastle, WJ.; Laver, J., editors. The Handbook of Phonetic Sciences. West Sussex, United Kingdom: Blackwell Publishing; 1999.

Hickok G, Poeppel D. The cortical organization of speech processing. Nat. Rev. Neurosci. 2007; 8:393–402. [PubMed: 17431404]

Kadia SC, Wang X. Spectral integration in A1 of a wake primates: neurons with single- and multipeaked tuning characteristics. J. Neurophysiol. 2003; 89:1603–1622. [PubMed: 12626629]

Kellis S, Miller K, Thomson K, Brown R, House P, Greger B. Decoding spoken words using local field potentials recorded from the cortical surface. J. Neural Eng. 2010; 7:056007. [PubMed: 20811093]

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychol. Rev. 1967; 74:431–461. [PubMed: 4170865]

Lofqvist, A. Theories and models of speech production. In: Hardcastle, WJ.; Laver, J., editors. The Handbook of Phonetic Sciences. West Sussex, United Kingdom: Blackwell Publishing; 1999.

Mesgarani N, Shamma S. Speech processing with a cortical representation of audio. Proc. ICASSP. 2011; 2011:5872–5875.

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

Mesgarani N, David SV, Fritz JB, Shamma SA. Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. J. Neurophysiol. 2009; 102:3329–3339. [PubMed: 19759321]

Miller LM, Escabi MA, Read HL, Schreiner CE. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. J. Neurophysiol. 2002; 87:516–527. [PubMed: 11784767]

Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA. Predicting human brain activity associated with the meanings of nouns. Science. 2008; 320:1191–1195. [PubMed: 18511683]

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. Bayesian reconstruction of natural images from human brain activity. Neuron. 2009; 63:902–915. [PubMed: 19778517]

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL. Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 2011; 21:1641–1646. [PubMed: 21945275]

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA 3RD, Brugge JF. Temporal envelope of time-compressed speech represented in the human auditory cortex. J. Neurosci. 2009; 29:15564–15574. [PubMed: 20007480]

Pasley BN, David SV, Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF. Reconstructing speech from human auditory cortex. PLoS Biol. 2012; 10(1):e1001251. http://dx.doi.org/10.1371/journal.pbio.1001251. Epub 2012 Jan 31. [PubMed: 22303281]

Pei X, et al. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. J. Neural Eng. 2011a; 8:046028. [PubMed: 21750369]

Pei X, Leuthardt EC, Gaona CM, Brunner P, Wolpaw JR, Schalk G. Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. Neuroimage. 2011b; 54:2960–2972. [PubMed: 21029784]

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 2009; 12:718–724. [PubMed: 19471271]

Rauschecker JP, Tian B, Pons T, Mishkin M. Serial and parallel processing in rhesus monkey auditory cortex. J. Comp. Neurol. 1997; 382:89–103. [PubMed: 9136813]

Robson H, Grube M, Lambon Ralph MA, Griffiths TD, Sage K. Fundamental deficits of auditory perception in Wernicke's aphasia. Cortex. 2013; 49:1808–1822. [PubMed: 23351849]

Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E. Sequential processing of lexical, grammatical, and phonological information within Broca's area. Science. 2009; 326:445–449. [PubMed: 19833971]

Schonwiesner M, Zatorre RJ. Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proc. Natl. Acad. Sci. U. S. A. 2009; 106:14611–14616. [PubMed: 19667199]

Schreiner CE, Read HL, Sutter ML. Modular organization of frequency integration in primary auditory cortex. Annu. Rev. Neurosci. 2000; 23:501–529. [PubMed: 10845073]

Sutter ML, Schreiner CE. Physiology and topography of neurons with multipeaked tuning curves in cat primary auditory cortex. J. Neurophysiol. 1991; 65:1207–1226. [PubMed: 1869913]

Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM. Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. J. Neurophysiol. 2004; 91:1282–1296. [PubMed: 14614108]

Tankus A, Fried I, Shoham S. Structured neuronal encoding and decoding of human speech features. Nat. Commun. 2012; 3:1015. [PubMed: 22910361]

Tankus A, Fried I, Shoham S. Cognitive-motor brain-machine interfaces. J. Physiol. Paris. 2013; (13) pii: S0928-4257(13)00035-1. http://dx.doi.org/10.1016/j.jphysparis.2013.05.005. [Epub ahead of print].

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. Network. 2001; 12:289–316. [PubMed: 11563531]

Todorov E. Optimality principles in sensorimotor control. Nat. Neurosci. 2004; 7:907–915. [PubMed: 15332089]

Viswanathan A, Freeman RD. Neurometabolic coupling in cerebral cortex reflects synaptic more than spiking activity. Nat. Neurosci. 2007; 10:1308–1312. [PubMed: 17828254]

Wrench, AA.; Hardcastle, WJ. Proceedings of the Fifth Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulator Modelling. Bavaria, Germany: Kloster Seeon; 2000. A multichannel articulatory speech database and its application for automatic speech recognition; p. 305-308.

Wu MC, David SV, Gallant JL. Complete functional characterization of sensory neurons by system identification. Annu. Rev. Neurosci. 2006; 29:477–505. [PubMed: 16776594]

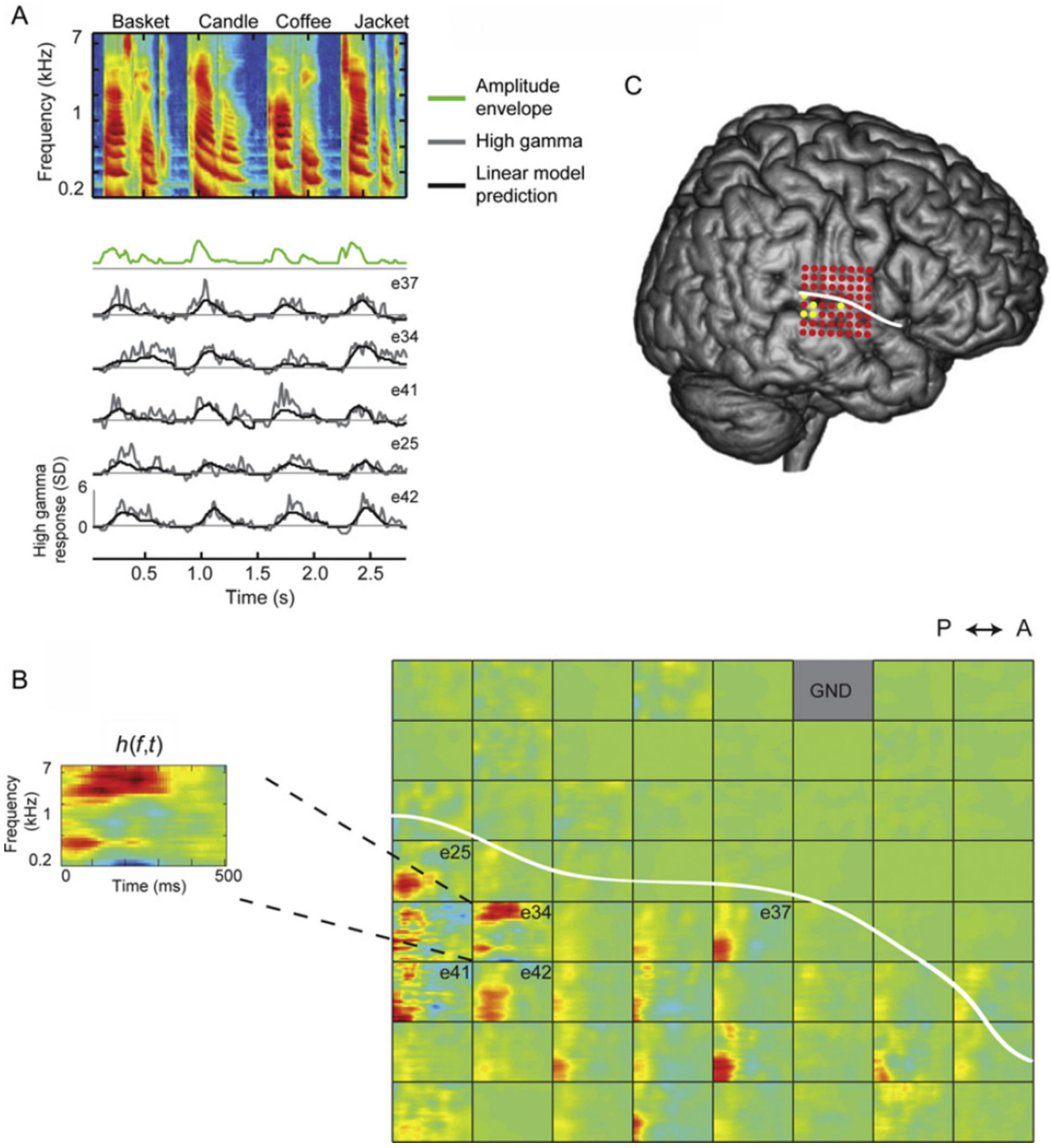

FIGURE 1.

(A) Example of single-trial ECoG responses in superior temporal gyrus (STG) to four spoken words. Top panel, spectrogram of four spoken words presented to the subject. Bottom panel, amplitude envelope of the speech stimuli (green), high-gamma ECoG neural responses at four different electrodes (gray), and predicted response from the spectrogram model (black). The ECoG responses are taken from five representative electrodes in STG (shown in yellow in C). (B) Spectrogram model, represented as $h(f, t)$, where h is the weight matrix as a function of frequency f and time t. This representation is equivalent to the

standard linear spectrotemporal receptive field (STRF). Positive weights (red) indicate stimulus components correlated with increased high-gamma activity, negative weights (blue) indicate components correlated with decreased activity, and nonsignificant weights (green) indicate no relationship. STRFs for each site in the electrode grid are shown (white curve marks the sylvian fissure). Anatomical distribution of these sites is shown in (C). Yellow circles indicate electrodes that are shown in (A).

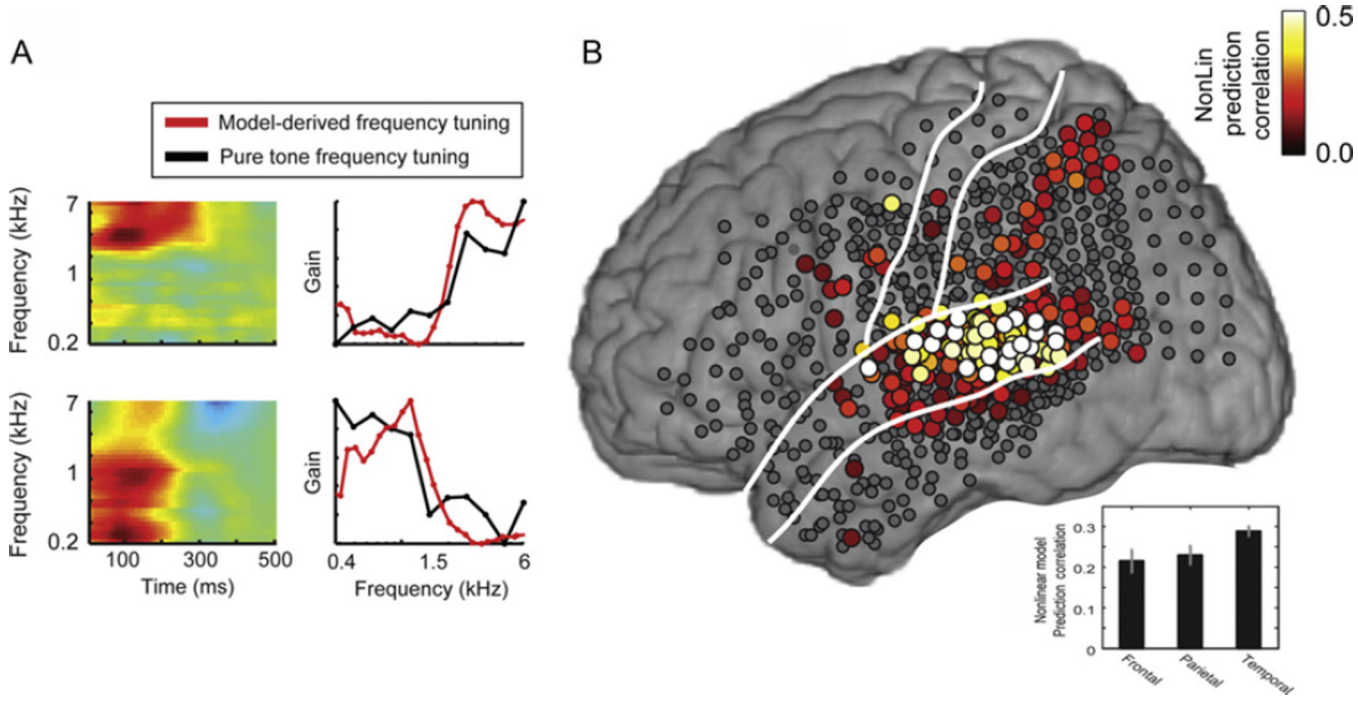

FIGURE 2.

(A) Fitted spectrogram models for 2 STG sites. Right panels; pure-tone frequency tuning (black curves) matches frequency tuning derived from fitted frequency models (red curves). Pure tones (375–6000 Hz, logarithmically spaced) were presented for 100 ms at 80 dB. Pure-tone tuning curves were calculated as the amplitudes of the evoked high-gamma response across tone frequencies. Model-derived tuning curves were calculated by first setting all inhibitory weights to zero and then summing across the time dimension (David et al., 2007). At these two sites, frequency tuning is either high-pass (top) or low-pass (bottom). (Reproduced from Pasley et al., 2012.) (b) Distribution of sites with significant modulation model predictive accuracy in the temporal, parietal, and frontal cortex.

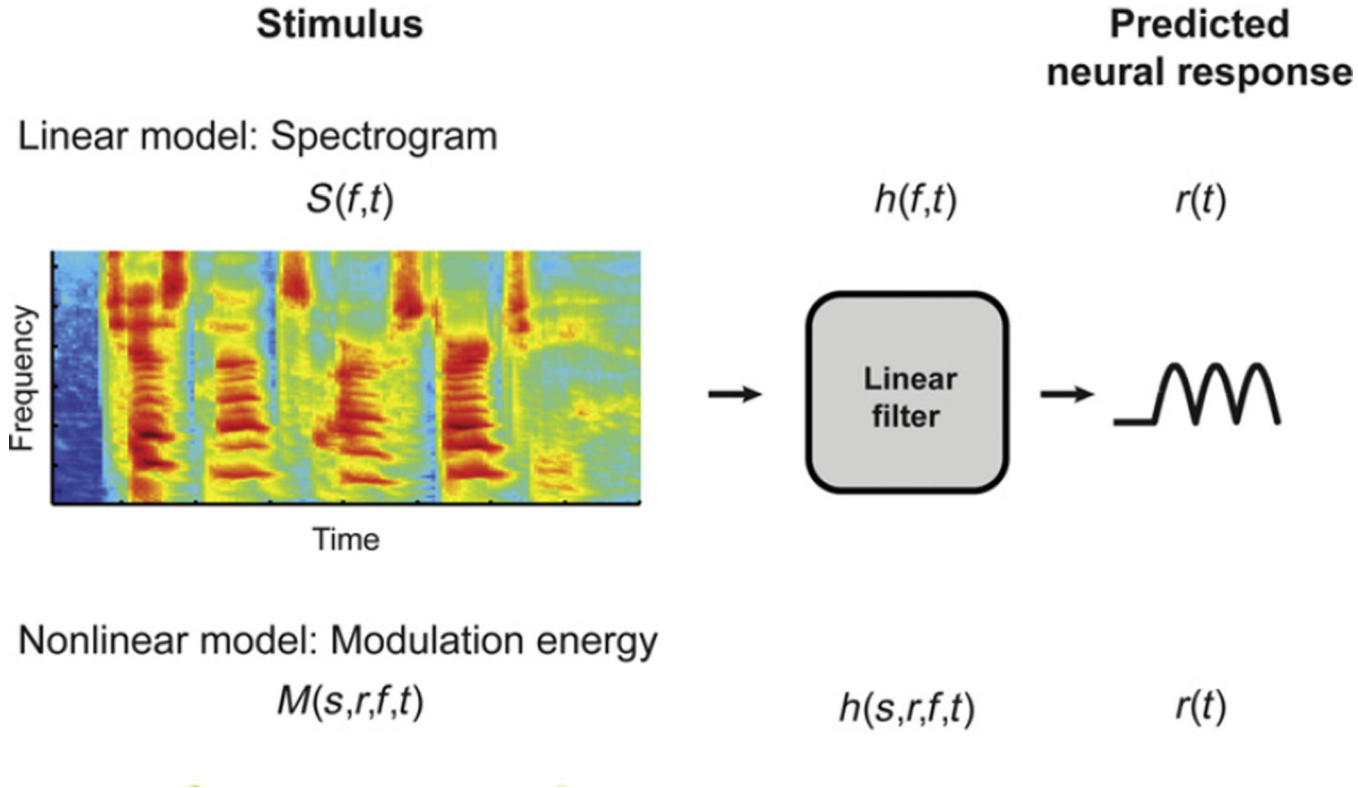

FIGURE 3.

Top panel, spectrogram model. The neural response across time $r(t)$ is modeled as a linear function $h(f, t)$ of the spectrogram representation of sound $S(f, t)$ where t is time, f is acoustic frequency, r is high-gamma neural activity, h is the weight matrix (STRF), and S is the acoustic spectrogram. For a single frequency channel, the instantaneous output may be high or low and does not directly indicate the modulation rate of the envelope. Bottom panel, modulation model. The neural response $r(t)$ is modeled as a linear function $h(s, r, f, t)$ of the modulation representation $M(s, r, f, t)$, where s is spectral modulation (scale) and r is temporal modulation (rate). The modulation encoding model explicitly estimates the modulation rate by taking on a constant value for a constant rate (Adelson and Bergen, 1985; Chi et al., 2005).

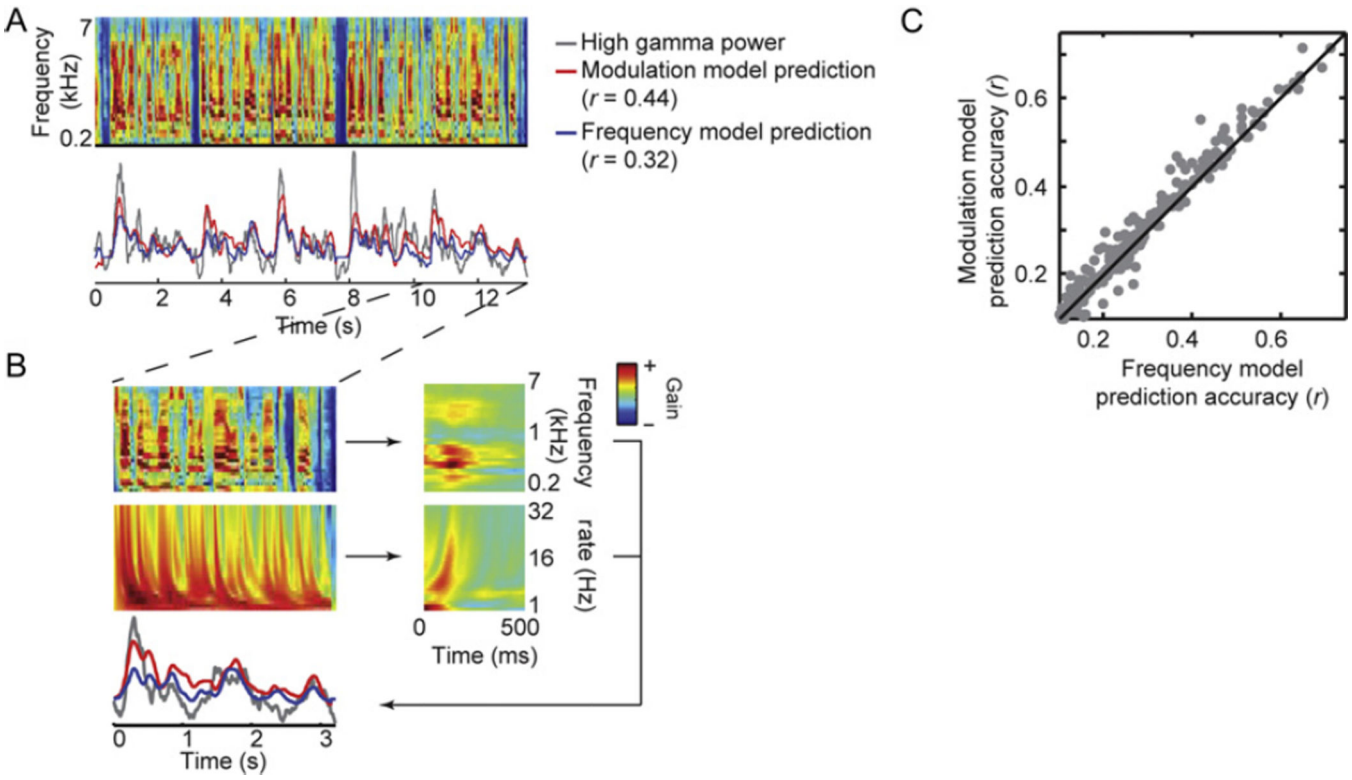

FIGURE 4.

(A) Example stimulus and response predictions from a representative electrode in the STG. High-gamma field potential responses (gray curve, bottom panel) evoked as the subject passively listened to a validation set of English sentences (spectrogram, top panel) not used in model fitting. Neural response predictions are shown for spectrogram (blue) and modulation models (red). The modulation model provides the highest prediction accuracy (r=0.44). (B) Example of fitted encoding models and response prediction procedure at an individual electrode site (same as in A). Top right panel; spectrogram model. Convolution of the STRF with the stimulus spectrogram generates a neural response prediction (bottom left panel, blue curve). Prediction accuracy is assessed by the correlation coefficient between the actual (bottom left panel, gray curve) and predicted responses. Bottom right panel; an example modulation energy model in the rate domain (for visualization, the parameters have been marginalized over frequency and scale axes). The energy model is convolved with the modulation energy stimulus representation (middle left panel) to generate a predicted neural response (bottom left panel, red curve). The energy and envelope models capture different aspects of the stimulus–response relationship and generate different response predictions. (C) Prediction accuracy of envelope versus modulation energy model across all predictive sites (n=199). The modulation energy model has higher prediction accuracy (p<0.005, paired t-test).

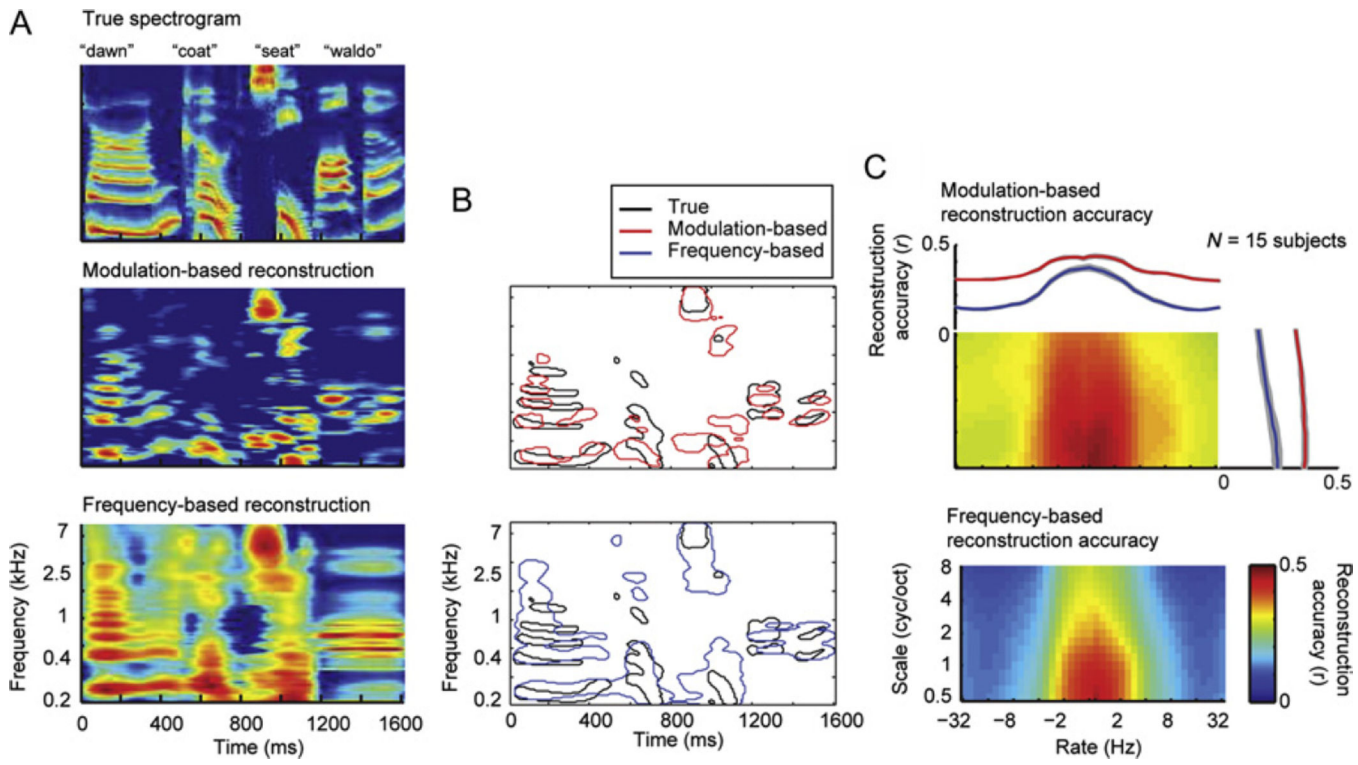

FIGURE 5.

(A) Top, the spectrogram of four English words presented aurally to the subject. Middle, the energy-based reconstruction of the same speech segment, which is linearly decoded from a set of responsive electrodes. Bottom, the envelope-based reconstruction, linearly decoded from the same set of electrodes. (B) The contours delineate the regions of 80% spectral power in the original spectrogram (black), energy-based reconstruction (top, red), and envelope-based reconstruction (bottom, blue). (C) Mean reconstruction accuracy (correlation coefficient) for the joint spectrotemporal modulation space across all subjects (N=15). Energy-based decoding accuracy is significantly higher compared to envelope-based decoding for temporal rates >2 Hz and spectral scales >2 cyc/oct (p<0.05, paired t-tests). Envelope decoding accuracy is maintained (r~0.3, p<0.05) for lower rates (<4 Hz rate, <4 cyc/oct scale), suggesting the possibility of a dual energy and envelope coding scheme for slower temporal modulations. Shaded gray regions indicate SEM (Pasley et al., 2012).

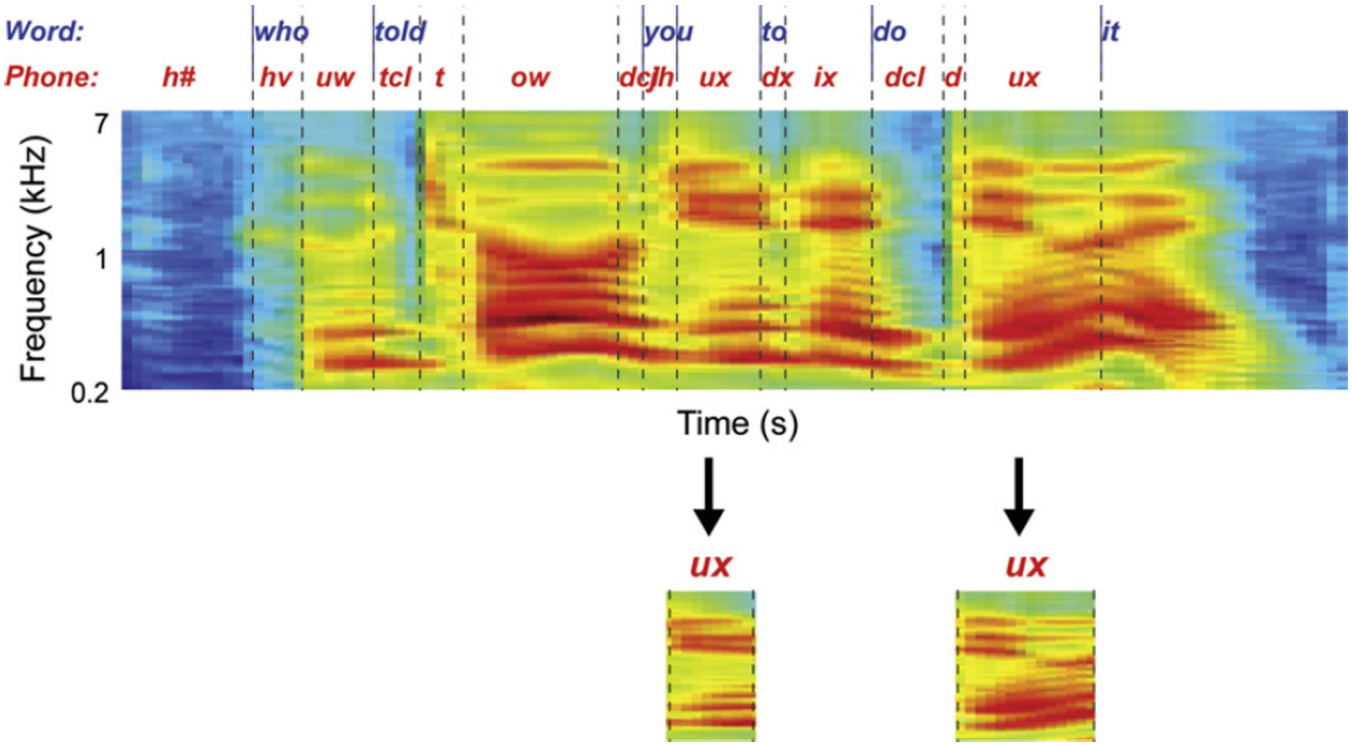

FIGURE 6.

The word and phonetic transcription of a sentence is shown. The vowel [ux] (TIMIT phonetic alphabet) occurs twice during the sentence. The spectrogram for the two instances differs as shown. The spectrogram encoding model assumes neural responses are sensitive to acoustic variation across phone instances. A phonetic model assumes neural responses are invariant to acoustic variability across phone instances.

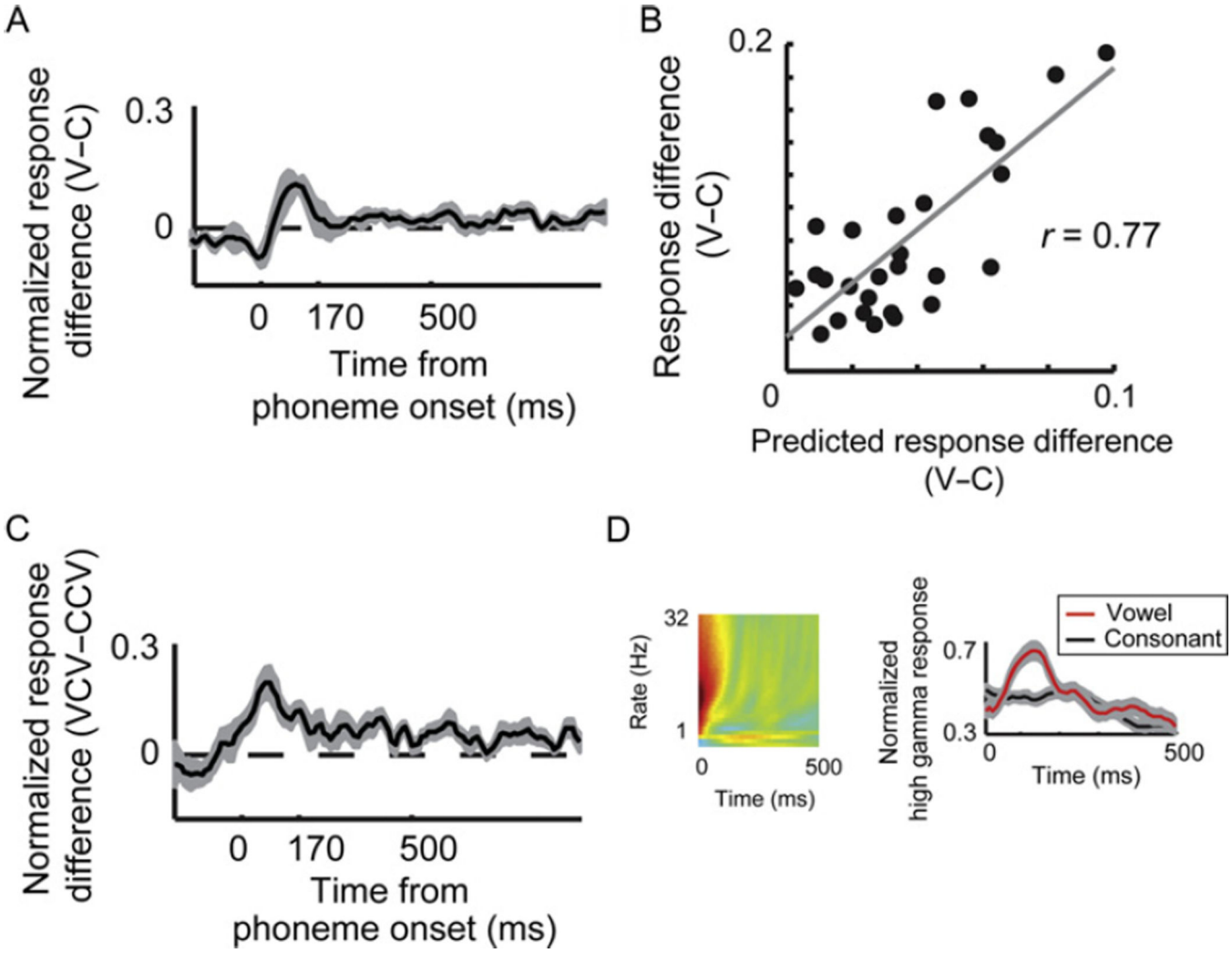

FIGURE 7.

Vowel-sensitive cortical sites and multisyllable responsivity. (A) The average high-gamma response difference (vowels, V, minus consonants, C) across all single syllable sites (n=5). Gray curves denote SEM over C/V occurrences. (B) The fitted energy models are used to filter a large set of English sentences and the average predicted response difference for consonants versus vowels is compared to the measured high-gamma response difference between the two classes. Across electrodes, the measured high-gamma CV response difference is highly correlated with that predicted from the energy model (r=0.77, $p<10^{-7}$). (C) The average high-gamma response difference (VCV–CCV) across all multisyllable sites (n=8). Time from phoneme onset is time-locked to the final vowel in the CCV or VCV sequence. (D) Left panel; example modulation model in the rate domain at a vowel-sensitive site. Right panel; average high-gamma response to consonants (C, blue curve) and vowels (V, red curve) embedded in English sentences. The high-gamma time series was first normalized by converting to z-scores. Gray curves denote SEM over CV occurrences.

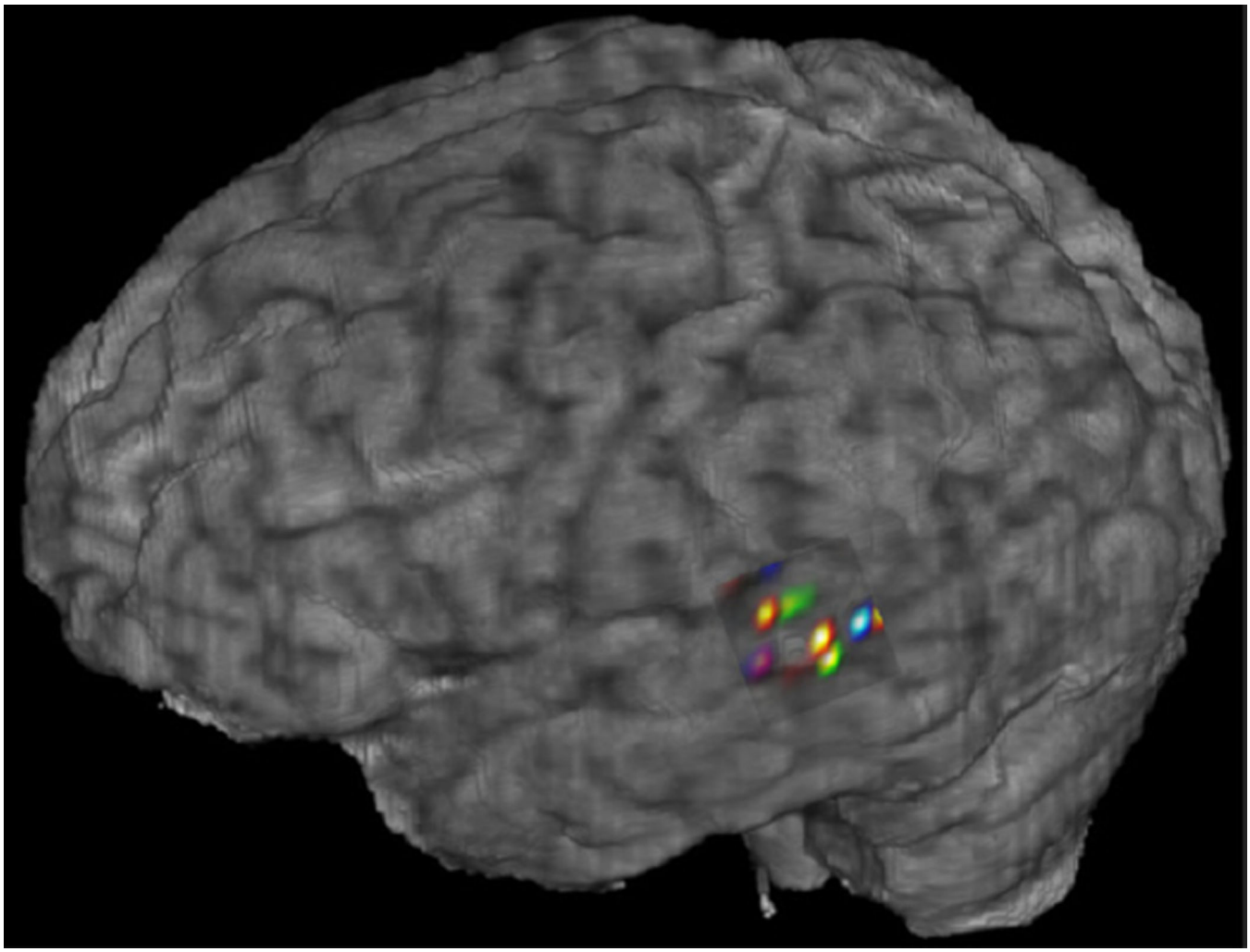

FIGURE 8.

Distribution of categorical responses to syllable perception in STG (Chang et al., 2010). Color indicates STG sites that discriminate specific pairs of syllables. Red: discriminates ba versus da; green: da versus ga; blue: ba versus ga. Mixed colors: electrode discriminates more than one pair. Phoneme decoding depends on distributed, interwoven networks with little overlap.

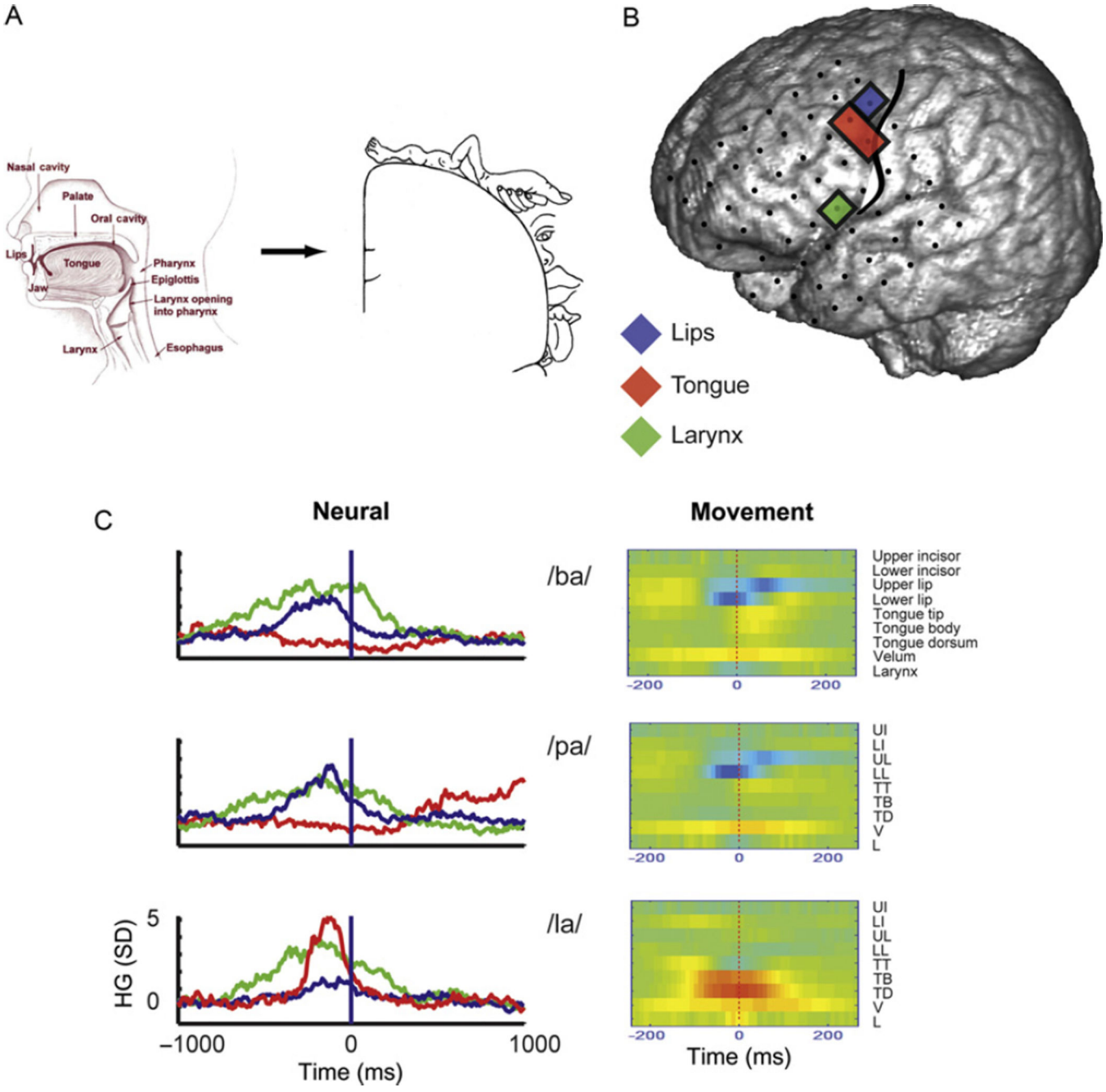

FIGURE 9.

Articulatory-based encoding model. (A) Upper panel, a hypothesized mapping of articulators to motor cortex. Muscles corresponding to various articulators in the vocal tract likely have anatomical representations in the motor homunculus. A “gestural score” (Browman and Goldstein, 1989) describes the temporal sequence of articulator activity during an utterance. The physical movement illustrated by the gestural score might then be “readout” via neural activity in the motor cortex. (B) Anatomical sites of three articulators in the motor map for a representative patient. Sites are determined both by electrical stimulation mapping performed during presurgical evaluation and by the presence of ECoG

activity during movement of individual articulators. (C) Left panel, high-gamma ECoG activity during the articulation of three CV monosyllables. Right panel, linear estimates of the articulator movement response (e.g., “gestural score”) for the same three consonants. The linear articulator response was derived from electromagnetic articulography measurements provided by the MOCHA speech corpus. Neural and articulator responses are qualitatively similar, indicating that motor map neural activity can be used to distinguish individual phonemes on the basis of articulatory patterns.