HHS Public Access

Author manuscript

Brain Lang. Author manuscript; available in PMC 2020 June 01.

Published in final edited form as:

Brain Lang. 2019 June ; 193: 73–83. doi:10.1016/j.bandl.2016.06.003.

The use of intracranial recordings to decode human language: challenges and opportunities

Stephanie Martin[1,2], **José del R. Millán**[1], **Robert T. Knight**[2,3], and **Brian N. Pasley**[2]

[1]Defitech Chair in Brain Machine Interface, Center for Neuroprosthetics, Ecole Polytechnique Fédérale de Lausanne, Switzerland [2]Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA [3]Department of Psychology, University of California, Berkeley, CA, USA

Abstract

Decoding speech from intracranial recordings serves two main purposes: understanding the neural correlates of speech processing and decoding speech features for targeting speech neuroprosthetic devices. Intracranial recordings have high spatial and temporal resolution, and thus offer a unique opportunity to investigate and decode the electrophysiological dynamics underlying speech processing. In this review article, we describe current approaches to decoding different features of speech perception and production – such as spectrotemporal, phonetic, phonotactic, semantic, and articulatory components – using intracranial recordings. A specific section is devoted to the decoding of imagined speech, and potential applications to speech prosthetic devices. We outline the challenges in decoding human language, as well as the opportunities in scientific and neuroengineering applications.

Keywords

intracranial recording; electrocorticography; speech decoding; spatio-temporal pattern of brain activity; time course; imagined speech; neuroprosthetics

1. Introduction

Language has been a topic of intense investigation for many decades, within many different disciplinary areas, including neurophysiological, psycholinguistic and behavioral studies. Early work in neurolinguistics was almost exclusively derived from brain lesion studies (Dronkers 1996; Watkins, Dronkers, and Vargha-Khadem 2002) and from non-invasive neuroimaging and electrophysiological data, such as functional magnetic resonance imaging, positron emission tomography (see Price 2012 for a review), magnetoencephalography and electroencephalography (see Ganushchak, Christoffels, and Schiller 2011 for a review).

Correspondence: Brian Pasley, University of California, Berkeley, Helen Wills Neuroscience Institute, Berkeley, CA 94720, USA, bpasley@berkeley.edu.

These approaches have defined fundamental language components of the human brain and their functional significance, but generally lack either the spatial or the temporal resolution to investigate rapid changes in the complex cortical network underlying speech processing (Canolty 2007). In contrast, intracranial recordings have emerged as a powerful tool to investigate higher cognitive function and the complex network of speech. In particular, this recording technique offers an enhanced view of the spatio-temporal aspects of neuronal populations supporting language (Lachaux et al. 2012).

Substantial efforts have aimed to develop new tools for analyzing these brain signals given the increasing amount of data recorded with intracranial electrode grids. Neural encoding models have attracted increasing interest in neuroscience, as they allow testing a hypothesis about the neural coding strategy under study (Wu, David, and Gallant 2006; Pasley et al., 2012; Mesgarani et al. 2014; Vu et al. 2011; Paninski, Pillow, and Lewi 2007). The encoding model identifies stimulus tuning properties of the neural response and alternative models can be compared using the model's predictive power (Wu, David, and Gallant 2006). Neural decoding models refer to the reverse mapping from brain response to stimulus, with the challenge being able to reconstruct sensory stimuli behavioral parameters from information encoded in the neural response. These models form the basis of neural interface systems that can be used to develop assistive technologies for people with disabling neurological conditions. For example, patients with long-standing tetraplegia were able to control a robotic arm to perform three-dimensional reach and grasp movements using a neural interface system (Hochberg et al. 2012)

Currently, more than two million people in the Unites States, and far more around the world, have verbal communication deficits, resulting from brain injury or disease (Wolpaw et al. 2002). People with such speech impairments would benefit from an internal speech decoder that can translate their brain activity in a natural and intuitive way. Until recently, efforts were mainly centered on motor and visual restoration. Fewer studies have investigated and decoded features of speech, in order to better understand the neural correlates of language for targeting of speech neuroengineering applications.

In this review article, we describe current approaches to decoding different features of speech perception and production, such as spectrotemporal, phonetic, semantic and articulatory components, using intracranial recordings. A specific section is devoted to the decoding of imagined speech. We also outline the challenges in decoding human language, as well as the opportunities and unique applications offered. To provide context, we first briefly describe the functional organization of language, as well as present the properties of intracranial recordings.

2. Functional organization of speech

Language is encoded in a widely distributed and complex network – whose activation depends on both linguistic context and brain modality. It is well accepted that two main brain areas involved in speech comprehension and speech production are Wernicke's area (posterior superior and middle temporal gyrus/superior temporal sulcus) and Broca's area

(posterior inferior frontal gyrus), respectively (Figure 1; see Price 2000; Démonet, Thierry, and Cardebat 2005 and Hickok and Poeppel 2007 for reviews).

Speech comprehension involves multiple stages of neural representations in order to convert sound to meaning. The first stage in this process involves spectrotemporal analysis of the acoustic signal in early auditory cortices. This is followed by phonetic and phonological processing in the superior temporal lobe (Hickok and Poeppel 2007), in which continuous acoustic features are projected into categorical representations (Chang et al. 2010). Ultimately, higher levels of speech comprehension transform intermediate speech representations into conceptual and semantic representations in the so-called ventral stream (superior middle temporal lobe).

In addition, the dorsal stream (posterior dorsal temporal lobe, parietal operculum and posterior frontal lobe) is responsible for translating speech signals into articulatory representations. The network projects from primary auditory cortices to more dorsal aspects of the temporal lobe, and then to the posterior frontal lobe (Broca's area), as well as premotor and supplementary areas (Hickok and Poeppel 2007). Broca's area, which originally was recognized to be important for speech motor production, has recently been challenged regarding its role. Recent work suggested that Broca's area coordinates speech – rather than actually providing the motor output – by mediating a cascade of activation from sensory representations to their corresponding articulatory gestures (Adeen Flinker et al. 2015). Speech production itself is the result of coordinated and precise movements over rapid time scales in the ventral half of the lateral sensorimotor cortex (Levelt 1993; Gracco and Löfqvist 1994; Brown et al. 2009), where the organization of the articulators (i.e., lips, jaw, tongue and larynx) is arranged somatotopically.

While actual speech perception and production have been extensively studied, the neural mechanisms underlying imagined speech remain poorly understood. Imagined speech (referred to as: inner speech, silent speech, speech imagery, covert speech or verbal thoughts) is defined here as the ability to actively generate internal speech representations, in the absence of any external speech stimulation or self-generated overt speech. Understanding and decoding the neural correlates of imagined speech has enormous potential for targeting speech devices. However, due to the lack of behavioral output and subjective nature of sensory imagery, imagined speech cannot be easily controlled and analyzed. It is well accepted that imagined speech share overlapping brain areas with both speech perception and production (Palmer et al. 2001; Aleman 2004; Geva et al. 2011). However, the different speech processing levels are also dissociated at the neural level (Huang, Carr, and Cao 2002; Shuster and Lemieux 2005; Leuthardt et al. 2012).

3. Intracranial recordings

Intracranial recording, also called electrocorticography (ECoG) has been used for decades in patients with epilepsy to localize the seizure onset zone, prior to brain tissue resection. In such cases, clinical procedure requires temporary implantation of electrode grids or strips onto the cortical surface, either above (epidural) or below (subdural) the dura mater (Figure 2). In some cases, depth electrodes (stereo electrodes) are implanted to identify brain

functions, and can have up to 3 mm inter-electrode spacing (Halgren, Marinkovic, and Chauvel 1998). In contrast, we refer to intracortical recordings when electrode shafts are inserted into deep brain structures and have both macroelectrodes along the electrode shaft and either splayed micro-wires at the recording tip or microwires along the shaft. This intracortical depth recording approach is referred to as steroencephalogpaphy (SEEG), and enables both field potential and single unit activity (SUA) recording from deep brain regions. Because of its invasiveness, intracranial recordings are applied exclusively for clinical purposes; nevertheless, the implantation time provides a unique opportunity to investigate human brain functions. Electrode grids placed over the temporal cortex, frontal cortex and sensory motor cortex are the most relevant for investigating speech processes.

ECoG has superb spatial (i.e., millimeter; Flinker et al. 2011) and spectral (0–500Hz; Staba et al. 2002) resolution, as well as higher amplitude (50–100μV) and signal-to-noise ratio – as compared to electroencephalography (EEG; centimeters, 0–40Hz, 10–20μV). It is also less sensitive to artifacts such as those generated by the electrical activity from skeletal muscle movements (Ball et al. 2009). In addition, the electrodes cover broad brain areas compared to intracortical recordings. Although, scalp EEG has a better overall brain coverage (e.g. covers both hemispheres), it has increased distortion and smearing of the electrical signal through to the skull, and therefore a much lower spatial resolution. Finally, ECoG has much higher temporal resolution (millisecond) with respect to metabolic imaging techniques, such as functional magnetic resonance imaging and positron emission tomography (seconds).

The high gamma frequency band (HG; 70–110 Hz) has been correlated with multiunit spike rate and asynchronous post-synaptic current of the underlying neuronal population (Manning et al. 2009; Lachaux et al. 2012; Buzsáki, Anastassiou, and Koch 2012). High gamma signal reliably tracks neural activity in many sensory modalities and correlates with cognitive functions, such as memory and speech (see Lachaux et al. 2012 for a review). The high gamma band has been correlated with spectrotemporal acoustic properties of speech in the superior temporal gyrus (Pasley et al. 2012; Kubanek et al. 2013), phonetic features (Chang et al. 2010), and articulatory features in the sensorimotor cortex (Bouchard et al. 2013). These studies demonstrate that many aspects of speech are robustly encoded in the high gamma frequency range of the ECoG signal.

Previous work has demonstrated the value of the ECoG signal in neuroprosthetic applications. For example, Leuthardt et al used ECoG recordings in humans to control a one-dimensional computer cursor (Leuthardt et al. 2004). Other ECoG-based brain-computer interfaces have been studied for the potential for assistive technologies (Schalk et al. 2007; Felton et al. 2007). Alternatively, formant frequencies (spectral peaks in the sound spectrum) were decoded in real time using intracortical recordings in the human motor cortex. The predicted speech was synthesized, and acoustically fed back to the user with a delay under 50ms (Guenther et al. 2009; Jonathan S. Brumberg et al. 2010).

Thus, ECoG represents a promising recording technique to investigate and decode neural correlates associated with human language. The next section briefly introduces neural encoding and decoding models and examples of applications to decode various speech features from the neural response.

4. Neural encoding and decoding models

A sensory stimulus, a movement, or a cognitive state generates a brain response that is specific in time, frequency band and location. Neural encoding models attempt to characterize how these parameters are represented in the brain. The model essentially asks the question, given a particular stimulus or behavior representation, can we predict the resulting brain response? On the other hand, a decoding model predicts information about the stimulus, behavior, or cognitive state from measured brain activity response, i.e., how well can we estimate the stimulus, behavior, or cognitive state given the brain activity? Various research fields have applied neural decoding models, including memory (Rissman, Greely, and Wagner 2010), vision (Kay and Gallant 2009) and motor research (Kubánek et al. 2009; Schalk et al. 2007). In neurorehabilitation, decoding models have allowed predicting 3D trajectories of a robotic arm for motor substitution (Hochberg et al. 2012).

Speech processing includes various processing steps – such as acoustic processing in the early auditory periphery, phonetic and categorical encoding in posterior areas of the temporal lobe and semantic and higher level of linguistic processes in later stages. One can ask what are the critical features of speech to target for efficient decoding and designing optimal communication technologies? For instance, a decoding model can synthetize acoustic features from parameters predicted from the brain activity. Alternatively, decoding discrete phonemes allows building words and sentences. Decoding speech is a complex problem, and can be approached with different strategies, objectives, and methods. Addressing the different models is central for our basic understanding of neural information processing, as well as for engineering speech neuroprosthetic devices. In this analytic framework, two main categories of decoding models have been studied (Figure 3); discrete category decoding and continuous feature decoding.

Discrete category decoding, also referred as classification, is the simplest form of decoding, in which the neural activity during specific events is identified from a finite set of possible events (Figure 3a). The classification accuracy often relies on sophisticated feature extraction techniques and classification algorithms developed in machine learning research (Varoquaux and Thirion 2014). Various speech features have been classified above chance levels, such as vowels and consonants during overt speech (Pei et al. 2011), phonemes (J.S. Brumberg et al. 2011; Mugler et al. 2014), syllables (Blakely et al. 2008), words (Kellis et al. 2010) and sentences (Dan Zhang et al. 2012). Decoding discrete units has been used in brain-computer interfaces to choose an action among a finite number of choices, and allow people control a wheelchair (Millán et al. 2009) or move a cursor on the screen (Wolpaw et al. 1991).

Continuous feature decoding, on the other hand, aims at reconstructing features of the stimulus under study (Figure 3b). Although relatively simple, techniques like linear regression can also be effective. For instance, the modeling of upper limb movement parameters, such as position, velocity and force has been extensively used to build motor prosthetic devices. Similarly, decoding models based on linear regression have been used to reconstruct visual (Nishimoto et al. 2011; or behavioral representations (Schalk et al. 2007) from brain activity. In speech reconstruction, acoustic features, such as formant frequencies

(Jonathan S. Brumberg et al. 2010), spectrotemporal fluctuations (Pasley et al. 2012) and mel-frequency cepstral-coefficients (Chakrabarti et al. 2013) were accurately reconstructed. In this approach, stimulus features are assumed to be a linear combination of the input (independent) variables – i.e., the brain activity. More complex hypotheses can be tested using linearizing decoding models, in which an intermediate feature space – a non-linear transformation of the stimulus representation – is linearly regressed against the neural activity (Naselaris et al. 2011). To evaluate the accuracy, the reconstructed stimuli are compared directly to the original representation. Reconstructing continuous features may allow synthesizing speech from the parameters decoded from brain activity. In a recent study, formant frequencies of intended speech were predicted directly from the activity of neurons involved in speech production. The predicted speech was synthesized from the decoded parameters and acoustically fed back to the user (Jonathan S. Brumberg et al. 2010). Such approaches have been extensively used in text-to-speech synthesis, and have recently been demonstrated to be very effective in synthesizing speech (King 2011).

In the next section, we describe the different speech representations that have been successfully reconstructed using electrocorticographic recordings.

4.1. Decoding spectrotemporal acoustic representations

An example of the types of speech features that can be targeted for decoding are spectro-temporal features of sound. Spectrotemporal features represent the spectrum of frequency in the sound waveform as they vary with time. A recent study demonstrated that spectrotemporal features could be reconstructed from high gamma frequency band ECoG signals (Pasley et al. 2012; Figure 4). In this study, two speech representations were evaluated for reconstruction: a spectrogram- and a modulation-based representation. The first representation was a time-varying representation of the amplitude envelope at each acoustic frequency – generated by an affine wavelet transformation of the sound waveform. This auditory filter was based on psychophysical and physiological studies of the cochlea, thus mimicking the frequency decomposition of sounds in the auditory periphery (Chi, Ru, and Shamma 2005). The second speech representation was based on a non-linear affine wavelet transformation of the spectrogram – reflecting temporal and spectral fluctuations in the spectrogram envelope. Spectral and temporal fluctuations carry essential phonological information that are robust under a variety of noise conditions, and reflect important properties of speech intelligibility (Chi et al. 1999; Elliott and Theunissen 2009). For instance, low and intermediate temporal modulation rates (<4Hz) are linked with syllable rate, whereas fast modulations (>16Hz) are related to syllable onsets and offsets. Similarly, broad spectral modulations are associated with vowel formants, whereas narrow spectral modulations are associated with harmonics (Shamma 2003).

Another continuous acoustic feature set that was successfully decoded is formant frequencies – which are concentrations of acoustic energy around particular frequencies in the speech waveform. A single case study recorded brain signals using intracortical depth electrodes in the human motor cortex and succeeded in predicting in real-time formant frequencies (Guenther et al. 2009). The audio signal was synthetized from the reconstructed acoustic features and fed back aurally to the patient within 50ms. This study highlights the

potential of using invasive neural recording techniques for building a neural-based speech synthesizer.

Altogether, these studies have shown that neural decoding models were able to successfully reconstruct continuous acoustic features of speech from early auditory cortices. The next encoding level in the speech processing stream is the mapping of continuous acoustic features into discrete phonological units – forming the building blocks of more complex speech utterances.

4.2. Decoding phonetic representations

Natural speech expression is not operated just under conscious control; it is also affected by various factors, including gender, emotional state, tempo, pronunciation and dialect. This leads to spectrotemporal speech irregularities both between speakers and within the same individual (Parneet Kaur and Vidushi Garg 2012). In addition, contextual effects and linguistic specifications also affect the acoustic realization of the speech sounds making up an utterance (King 2011) – i.e., preceding/following phonemes (co-articulation), position of segment in syllable, and end tone of phrase. As such, speech perception requires the rapid and effortless extraction of meaningful and invariant phonetic information from a highly variable acoustic signal. As an example, the phenomenon of categorical speech perception, where a continuum of acoustically varying sounds is perceived as perceptually distinct phonetic objects, was found to be encoded in the superior temporal gyrus (Chang et al. 2010).

From a decoding perspective, several studies have also shown successful classification of individual speech units into different categories. For instance, individual vowels and consonants during overt and covert speech production were classified above chance level (Pei et al. 2011). Similarly, phonemes were accurately predicted from intracranial recordings from the motor cortex (Mugler et al. 2014). In these studies, the discriminant information was extracted from various anatomical brain areas, and reflected different levels of speech processing, such as early spectrotemporal acoustic features, phonetic or articulatory movement components.

4.3. Decoding phonotactic sequences

Speech units are often analyzed as isolated components, whereas in natural speech communication, they are rarely found in isolation. Instead speech units are embedded in streams of natural speech where they are arranged sequentially (phonotactics; Vitevitch et al. 1999 to convey more complex linguistic and semantic meanings.

A recent study by Herff and colleagues showed that continuous spoken speech could be decoded into textual representations of single phone sequences, using Gaussian models as generative statistical representation for broadband gamma power (Herff et al. 2015, see Figure 5). During the decoding procedure, the information about the observed neural activity was combined with statistical language information – in order to find the most likely sequence of phones and words, given the observed neural activity sequence.

The model used in this study was context-independent (i.e., only one model trained for each phone, with no consideration of preceding or succeeding phones), whereas in auditory perception, listeners are normally tuned to the statistical properties of surrounding units (Winkler, Denham, and Nelken 2009). For instance, as reported by Leonard and colleagues, hearing the sound /k/ followed by /uw/ (“koo”) is more common than hearing /k/ followed by /iy/ (“kee”). Thus, in English, /k/ predicts /uw/ more strongly than /iy/, which is less likely to be the next sound (Leonard et al. 2015). Behavioral studies have shown that people learn these statistics, although they are not consciously aware of the probability distributions. A recent study investigated how the brain encodes the statistical structure of sequentially arranged unit (Leonard et al. 2015). Results showed that the neural response is modulated according to the language-level probability of surrounding sounds. In addition, phonotactic statistics integrate lower-level acoustic features in a context-dependent manner and is influenced by higher-level lexical features.

4.4. Decoding higher levels of speech processing

Speech processing in higher brain areas involves the integration of semantic meaning. A recent study showed that semantic information could be decoded from human intracranial recordings of the inferior frontal gyrus and superior temporal gyrus (Wei Wang et al. 2011). Additionally, natural verbal communication often takes place in noisy environments and background speech. As such, speech requires additional verbal working memory (Conway, Cowan, and Bunting 2001) and attentional resources (Fritz et al. 2007) to segregate the target stream from competing background noise. As an example of higher level precessing step, speech spectrograms of attended speech was succesfully decoded from neural activity associated with mixtures of speakers (Mesgarani and Chang 2012).

4.5. Decoding articulatory representations

Speech production is believed to occur in various processing levels, such as conceptual preparation, lexical selection, phonological code retrieval and encoding, phonetic encoding and articulation (Levelt, Roelofs, and Meyer 1999; Indefrey and Levelt 2004). The articulatory mechanisms involve various steps to select specific articulator muscles, identify the degree of activation of each muscle, and initiate a coordinated activation sequence (Levelt 1993).

A recent study provided evidence for how speech articulators are organized in the ventral sensorimotor cortex (vSMC; Bouchard et al. 2013). The decoding weights for the different articulators (lips, tongue, larynx and jaws) were somatotopically distributed across the vSMC, yet partially overlapping at individual electrode. Articulators were temporally coordinated as the production of consonant-vowels syllable unfolded sequentially, and were separable according to the phonetic features. This suggests that neural populations in the sensorimotor cortex represent both low-level parameters of movements (i.e., muscle activation) and high-level aspects (i.e., as movement goals and phonetic representation).

The studies described in this review article shows that both the low- and high-level parameters of speech processing might be targeted for decoding, and building speech prosthetic devices.

4.6. Decoding imagined speech representations

Imagined speech has been studied extensively (see Price 2012 and Perrone-Bertolotti et al. 2014 for reviews) yet the underlying neural representation remains poorly understood. This is mainly due to the subjective nature of imagined speech, which prevents measurement of behavioral output. For the same reason, it is difficult to build decoding models that directly regress the neural activity to any behavioral metric or speech representation. To address that issue, several approaches have been proposed.

A number of studies have provided evidence for decoding of neural activity associated with imagined speech features into categorical representations – i.e., covertly articulated isolated vowels (Ikeda et al. 2014), vowels and consonants during covert word production (Pei et al. 2011) and intended phonemes (J.S. Brumberg et al. 2011). These studies used spectral features to predict the class index among a finite number of choices, thus reducing the problem associated with speech production temporal irregularities.

An alternative possibility is to exploit the high temporal resolution offered by ECoG, and use time features for decoding. Recently, we evaluated the possibility to reconstruct continuous spectrotemporal acoustic features of imagined speech from a decoder built during overt speech production. This strategy was based on evidence that speaking out loud and speaking covertly may share common neural mechanisms (Palmer et al. 2001; Aleman 2004; Geva et al. 2011). In this study, the decoding model was built based on high gamma signals to reconstruct spectrotemporal auditory features of self-generated overt speech (Figure 6.a). Then, the same decoding model was applied to reconstruct auditory speech features in the covert speech condition (Figure 6.b). To evaluate performances, the reconstruction in the imagined speech condition was compared to the representation of the corresponding original sound spoken out loud – using a temporal realignment algorithm. Results showed that significant acoustic features of imagined speech could be reconstructed from models that were built from overt speech data (Martin et al. 2014). This supported the hypothesis that overt and covert speech share underlying neural mechanisms. In addition, the ability to decode imagined speech may provide a basis for development of a brain-based communication method for patients with disabling neurological conditions. Decoding speech – and imagined speech in particular – presents several distinct challenges as reviewed in the next section.

5. Challenges

5.1. Experimental barriers

Animal models have been extensively studied in most sensory and cognitive domains, providing fundamental descriptions of underlying neural mechanisms. However, speech is exclusive to human beings in comparison to other forms of communication used by non-human animals. Animals use a system of communication that is believed to be limited to expression of a finite number of utterances that is mostly determined genetically (Tomasello 2008). In contrast, humans can produce a vast range of utterances from a finite set of elements (Trask 1999). Lower-level auditory and motor processing has been widely explored in nonhuman mammals (Georgopoulos, Kettner, and Schwartz 1988; deCharms 1998;

Depireux et al. 2001) and avians (de Boer 1967; Theunissen et al. 2001). However, knowledge about physiological correlates in higher levels of speech processing likely cannot be inferred from animal models.

In addition, recordings in humans are generally restricted to noninvasive techniques such as EEG, MEG or fMRI. These approaches give large-scale overviews of cortical activity in distributed language networks, but typically lack either spatial or temporal resolution. A few intracortical recordings have shown promising results in decoding intended phonemes (J.S. Brumberg et al. 2011) and formant frequencies (Guenther et al. 2009). Although intracortical recordings obtained with stereo EEG (SEEG) have higher spatial resolution than cortical based intracranial recordings, their spatial coverage is limited – making them less suitable to investigate higher speech processing levels. Given its unique spatiotemporal properties, intracranial recording is a good candidate to decode speech, but its opportunities are limited in humans. Only in rare cases, patients with epilepsy undergoing neurosurgical procedure for brain ablation are implanted with ECoG grids. The ECoG grids provide the opportunity to investigate neuroanatomical pathways of language processing, but the configuration, location and duration of implantation are not designed for the experiments, but rather solely for clinical purposes. In addition, long-term implantation abilities in human is lacking, as compared to non-human primate studies that showed stable neural decoding for extended periods of time (weeks to months; Ashmore et al. 2012).

Long-term grid implantations in humans are desired for targeting neural prosthetic applications, but until now they still are limited by technical difficulties. One key issue is the foreign body response and increased impedance leading to loss of signal (Groothuis et al. 2014). This is particularly problematic for SEEG intracortical electrodes. Device material and electrode-architecture influences the tissue reaction. Softer neural implants with shape and elasticity of dura mater increase electrode conductivity and improve the implant-tissue integration (Minev et al. 2015). The design of the intracranial recording electrodes has been shown to be an important factor in motor decoding performance. Namely, the spatial resolution of a cortical surface electrode array depends on the size and spacing of the electrodes, as well as the volume of tissue to which each electrode is sensitive (Wodlinger et al. 2011). Many researchers have attempted to define what could be the optimal electrode spacing and size (Slutzky et al. 2010), but this is still an open area of research. There is emerging evidence from the brain computer interface literature that decoding performances might be improved when high gamma activity is derived from very high-density grids (Blakely et al. 2008; Rouse et al. 2013). However, although a smaller inter-electrodes spacing increases the spatial resolution, it poses additional technical issues related to the electrode grid design. This is a fundamental issue for speech decoding, given that speech processing at the individual neuron level revealed a complex pattern of individual cell firing (Chan et al. 2014).

6.2. Lack of understanding of language organization

As a result of such experimental barriers, the complex neural mechanisms underlying speech and language remain largely unknown. In addition, these functions are highly dependent on

the context and brain modality. Understanding speech processing is a key step to building efficient natural speech prosthesis.

Context refers to the factors that affect the acoustic realization of speech sounds – including segmental elements, such as co-articulatory features, and supra-segmental elements, such as stress, prosodic patterns, phonation type, and intonation. While context dependent modeling is very common in speech recognition (Waibel and Lee 1990) and known to significantly improve recognition performances, it has rarely been taken into account for neural decoding. A reason for this is that it remains unknown how the brain encodes the various factors affecting the production of speech sound. A key aspect for improving speech prosthetic will be to determine which factors significantly improve decoding performances, and how to model them.

Language is involved in a variety of modalities including writing, reading, listening and speaking. These four modalities share receptive and production areas of the brain, yet, they also have unique processing levels and neuroanatomical substrates (Berninger and Abbott 2010; Singleton and Shulman 2014). For instance, a person with a writing deficit may still speak normally, or vice versa (Rapp, Fischer-Baum, and Miozzo 2015). The locations of brain injuries are known to lead to different types of language deficits. Various types of aphasia that affect language components range from auditory, phonological or lexical functions (Pasley and Knight 2013). This complex network connects expressive and receptive brain areas, processing oral and written forms of language.

To better understand the effect of context and modality of speech decoding, efforts should be made to distinguish what features of speech are encoded in the different frequency bands. Most of the studies described in this review have focused on the decoding of HG frequency bands, but there is evidence that lower frequency bands encode various aspects of speech, such as attended speech envelope (Ding and Simon 2012; Zion Golumbic et al. 2013), and track and discriminate spoken sentences (Luo and Poeppel 2007). As a result, lower frequency bands might provide complementary information to the more traditional HG frequency band, and overcome the shortcomings of some of other frequency bands.

Finally, neural activity associated with speech is not stationary, but modulated by top-down influences based on expectations (Leonard et al. 2015), and speech monitoring feedback loops (Chang et al. 2013; Houde and Chang 2015). As an example, speakers are known to articulate differently when deprived of auditory feedback of their own speech, such as in high-noise environments. This is an additional challenge that will have to be faced to decode speech in natural settings.

5.3. Modeling limitations

A major challenge in neural decoding applications lies in the computational models used for decoding the neural correlates of language. Most of the studies described in this review use linear decoding models to map the neural activity to the speech representations. However, in reality, the neural correlates of language are likely non-linearly related to the various speech representations.

As speech sounds unfold, brain signals track spectrotemporal acoustic features, as well as phonetic and phonotactic elements. One question arises from these complex processing steps: what is the best speech representation to be decoded and how to model it. Answering this question is central for our basic understanding of neural information processing, as well as for engineering speech neuroprosthetic devices (Donoghue 2002). As mentioned earlier, two large categories of decoders can be targeted: discrete and continuous modeling approaches. Both categories pose unique set of challenges.

In discrete decoding approaches, the neural activity is classified into a class among a finite number of choices based on similarities. Studies in speech recognition/synthesis have investigated for over eighty years the optimal speech unit size to be analyzed; the answer remains a matter of debate. However, the longer the unit, the larger the database needed to cover the required domain, while smaller units offer more degrees of freedom, and can build a larger set of complex utterances. Alternatively, decoding individual words carries in itself more semantic information, which would be relevant in a basic clinical setting; e.g., decode one among ten clinically relevant words (‘hungry’, ‘thirsty’, ‘yes’, ‘no’, etc.). A challenge will be the creation of dictionaries of limited size, but that are rich enough to be useful for speech communication in neurological patients.

In continuous decoding approaches, the goal is not to predict the label of a trial, but to reconstruct continuous features. A difficulty in this approach lies in defining what are the best speech parameters to model. Potential avenues for alternative speech-neuroprosthetics would be to synthetize audible and understandable speech directly from neural decoding.

6. Opportunities

Neural decoding models have attracted increasing interest as novel research tools to derive data driven hypotheses underlying complex cognitive functions. Progress on uncovering the link between speech and neural responses revealed that brain activity is tuned to various levels of speech descriptions, broken down into anatomic and functional stages. Decoding speech from intracranial recordings serves two complementary purposes: understanding the neural correlates of speech processing and decoding speech features for targeting speech neuroprosthetic devices. For instance, understanding how speech is processed in the brain leads to better design and implementation of robust and efficient models for decoding speech representations. Alternatively, decoding models allow validating hypotheses about the neural coding strategy under study.

Unique opportunities for targeting speech neuroprosthesis are offered by combining different research fields. First, research findings from neuroscience reveal which anatomical locations and brain signals should be modeled. Second, linguistic fields support development of decoding models that incorporate linguistic, contextual specifications – including segmental elements and supra-segmental elements. Combining insights from these research fields with machine learning and statistical modeling is a key element to improve prediction accuracies. Finally, the success of speech neuroprosthesis will depend on the continuous technological improvements to enhance signal quality and resolution, and allow developing more portable and biocompatible invasive recording devices.

The knowledge coming from these various scientific fields will help to build hybrid-decoding systems, and provide realistic applications for people with speech production impairments. Hybrid systems acquire sensor data from multiple elements of the human speech production system, and combines the different signals to optimize speech synthesis (see Brumberg et al. 2010 for a review). For instance, recording sensors allow characterizing the vocal tract by measuring its configuration directly or by sounding it acoustically using electromagnetic articulography, ultra-sound or optical imaging of the tongue and lip. Alternatively, electrical measurements can infer articulation from actuator muscle signals (i.e., using surface electromyography) or signals obtained directly from the brain (mainly EEG and ECoG). Using different sensors and different speech representations allow exploiting an individual's residual speech functions to operate the speech synthesis.

The various types of language deficits exemplify the challenge in targeting specific speech prosthesis that addresses individual needs. As a first step, encoding models offers a possible functional explanation for specific language disorder, and identify injured neural circuits. For instance, a recent study identified impairments in basic spectrotemporal modulation processing of auditory stimuli in Wernicke's aphasia patients with lesions to parietal and superior temporal areas (Robson et al. 2013). Encoding modeling approaches also offers the opportunity to measure continuously changes in cortical representation induced by rehabilitation. Quantitative measures of plasticity would allow optimizing and guiding training-induced changes in specific cortical areas, and applicable to a variety of aphasic symptoms having different level of speech representation affected.

Both receptive and expressive types of aphasia could benefit from encoding models and targeted rehabilitation. In addition, expressive language impairments could also benefit from decoding models. Especially, when the motor output is disrupted, but the internal speech representations remain intact. In these cases, speech devices could assist and communicate out loud what they cannot express anymore – using various speech features along the different processing levels. As such challenges are solved, decoding speech opens the door to new communication interfaces that may allow for more natural speech-like communication to take place when no audible acoustic signal is available in patients with severe communication deficits.

Acknowledgments

This work was supported by the Zeno-Karl Schindler Foundation, NINDS Grant R3721135, K99DC012804, and the Nielsen Corporation.

References

Agosta, Federica; Henry, Roland G; Migliaccio, Raffaella; Neuhaus, John; Miller, Bruce L; Dronkers, Nina F; Brambati, Simona M; , et al. Language Networks in Semantic Dementia. Brain: A Journal of Neurology. 2010; 133(Pt 1):286–299. [PubMed: 19759202]

Aleman A. The Functional Neuroanatomy of Metrical Stress Evaluation of Perceived and Imagined Spoken Words. Cerebral Cortex. 2004; 15(2):221–228. [PubMed: 15269107]

Ashmore, RC, Endler, BM, Smalianchuk, I, Degenhart, AD, Hatsopoulos, NG, Tyler-Kabara, EC, Batista, AP, Wang, W. Stable Online Control of an Electrocorticographic Brain-Computer Interface Using a Static Decoder. IEEE; 2012. 1740–1744.

Ball, Tonio; Kern, Markus; Mutschler, Isabella; Aertsen, Ad; Schulze-Bonhage, Andreas. Signal Quality of Simultaneously Recorded Invasive and Non-Invasive EEG. NeuroImage. 2009; 46(3): 708–716. [PubMed: 19264143]

Berninger, Virginia W; Abbott, Robert D. Listening Comprehension, Oral Expression, Reading Comprehension, and Written Expression: Related yet Unique Language Systems in Grades 1, 3, 5, and 7. Journal of Educational Psychology. 2010; 102:635–651. [PubMed: 21461140]

Blakely, Timothy; Miller, Kai J; Rao, Rajesh PN; Holmes, Mark D; Ojemann, Jeffrey G. Localization and Classification of Phonemes Using High Spatial Resolution Electrocorticography (ECoG) Grids; Conference Proceedings: … Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference; 2008. 4964–4967. 2008

Bouchard, Kristofer E; Mesgarani, Nima; Johnson, Keith; Chang, Edward F. Functional Organization of Human Sensorimotor Cortex for Speech Articulation. Nature. 2013; 495(7441):327–332. [PubMed: 23426266]

Brown, Steven; Laird, Angela R; Pfordresher, Peter Q; Thelen, Sarah M; Turkeltaub, Peter; Liotti, Mario. The Somatotopy of Speech: Phonation and Articulation in the Human Motor Cortex. Brain and Cognition. 2009; 70(1):31–41. [PubMed: 19162389]

Brumberg, Jonathan S; Nieto-Castanon, Alfonso; Kennedy, Philip R; Guenther, Frank H. Brain–computer Interfaces for Speech Communication. Speech Communication. 2010; 52(4):367–379. [PubMed: 20204164]

Brumberg JS, Wright EJ, Andreasen DS, Guenther FH, Kennedy PR. Classification of Intended Phoneme Production from Chronic Intracortical Microelectrode Recordings in Speech-Motor Cortex. Frontiers in Neuroscience. 2011 May.

Buzsáki, György; Anastassiou, Costas A; Koch, Christof. The Origin of Extracellular Fields and Currents — EEG, ECoG, LFP and Spikes. Nature Reviews Neuroscience. 2012; 13(6):407–420. [PubMed: 22595786]

Canolty, Ryan T. Spatiotemporal Dynamics of Word Processing in the Human Brain. Frontiers in Neuroscience. 2007; 1(1):185–196. [PubMed: 18982128]

Chakrabarti, S; Krusienski, Dean J; Schalk, Gerwin; Brumberg, Jonathan S. Predicting Mel-Frequency Cepstral Coefficients from Electrocorticographic Signals during Continuous Speech Production; 6th International IEEE/EMBS Conference on Neural Engineering (NER); 2013. http://neuro.embs.org/files/2013/0607_FI.pdf

Chan AM, Dykstra AR, Jayaram V, Leonard MK, Travis KE, Gygi B, Baker JM, et al. Speech-Specific Tuning of Neurons in Human Superior Temporal Gyrus. Cerebral Cortex. 2014; 24(10):2679–2693. [PubMed: 23680841]

Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF. Human Cortical Sensorimotor Network Underlying Feedback Control of Vocal Pitch. Proceedings of the National Academy of Sciences. 2013; 110(7):2653–2658.

ChangRieger, Jochem W; Johnson, Keith; Berger, Mitchel S; Barbaro, Nicholas M; Knight, Robert T. Categorical Speech Representation in Human Superior Temporal Gyrus. Nature Neuroscience. 2010; 13(11):1428–1432. [PubMed: 20890293]

Chi T, Gao Y, Guyton MC, Ru P, Shamma S. Spectro-Temporal Modulation Transfer Functions and Speech Intelligibility. The Journal of the Acoustical Society of America. 1999; 106(5):2719–2732. [PubMed: 10573888]

Chi T, Ru P, Shamma SA. Multiresolution Spectrotemporal Analysis of Complex Sounds. The Journal of the Acoustical Society of America. 2005; 118(2):887. [PubMed: 16158645]

Conway, Andrew RA; Cowan, Nelson; Bunting, Michael F. The Cocktail Party Phenomenon Revisited: The Importance of Working Memory Capacity. Psychonomic Bulletin & Review. 2001; 8(2):331–335. [PubMed: 11495122]

Dan, Zhang; Gong, Enhao; Wu, Wei; Lin, Jiuluan; Zhou, Wenjing; Hong, Bo. Spoken Sentences Decoding Based on Intracranial High Gamma Response Using Dynamic Time Warping. IEEE; 2012. 3292–3295.

de Boer E. Correlation Studies Applied to the Frequency Resolution of the Cochlea. Journal of Auditory Research. 1967; 7

deCharms RC. Optimizing Sound Features for Cortical Neurons. Science. 1998; 280(5368):1439–1444. [PubMed: 9603734]

Démonet, Jean-François; Thierry, Guillaume; Cardebat, Dominique. Renewal of the Neurophysiology of Language: Functional Neuroimaging. Physiological Reviews. 2005; 85(1):49–95. [PubMed: 15618478]

Depireux DA, Simon JZ, Klein DJ, Shamma SA. Spectro-Temporal Response Field Characterization with Dynamic Ripples in Ferret Primary Auditory Cortex. Journal of Neurophysiology. 2001; 85(3):1220–1234. [PubMed: 11247991]

Ding N, Simon JZ. Emergence of Neural Encoding of Auditory Objects While Listening to Competing Speakers. Proceedings of the National Academy of Sciences. 2012; 109(29):11854–11859.

Donoghue, John P. Connecting Cortex to Machines: Recent Advances in Brain Interfaces. Nature Neuroscience. 2002 Nov; 5(Suppl):1085–1088. [PubMed: 12403992]

Dronkers NF. A New Brain Region for Coordinating Speech Articulation. Nature. 1996; 384(6605): 159–161. [PubMed: 8906789]

Elliott, Taffeta M; Theunissen, Frédéric E. The Modulation Transfer Function for Speech Intelligibility. PLoS Computational Biology. 2009; 5(3):e1000302. [PubMed: 19266016]

Felton, Elizabeth A; Wilson, J. Adam; Williams, Justin C; Garell, P. Charles Electrocorticographically Controlled Brain-Computer Interfaces Using Motor and Sensory Imagery in Patients with Temporary Subdural Electrode Implants. Report of Four Cases. Journal of Neurosurgery. 2007; 106(3):495–500.

Flinker A, Chang EF, Barbaro NM, Berger MS, Knight RT. Sub-Centimeter Language Organization in the Human Temporal Lobe. Brain and Language. 2011; 117(3):103–109. [PubMed: 20961611]

Flinker, Adeen; Korzeniewska, Anna; Shestyuk, Avgusta Y; Franaszczuk, Piotr J; Dronkers, Nina F; Knight, Robert T; Crone, Nathan E. Redefining the Role of Broca's Area in Speech. Proceedings of the National Academy of Sciences. 2015; 112(9):2871–2875.

Fritz, Jonathan B; Elhilali, Mounya; David, Stephen V; Shamma, Shihab A. Auditory Attention—focusing the Searchlight on Sound. Current Opinion in Neurobiology. 2007; 17(4):437–455. [PubMed: 17714933]

Ganushchak, Lesya Y; Christoffels, Ingrid K; Schiller, Niels O. The Use of Electroencephalography in Language Production Research: A Review. Frontiers in Psychology. 2011; 2

Georgopoulos AP, Kettner RE, Schwartz AB. Primate Motor Cortex and Free Arm Movements to Visual Targets in Three-Dimensional Space. II. Coding of the Direction of Movement by a Neuronal Population. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 1988; 8(8):2928–2937. [PubMed: 3411362]

Geva, Jones; Baron, Crinion; Warburton, E. The Neural Correlates of Inner Speech Defined by Voxel-Based Lesion-Symptom Mapping. Brain. 2011; 134(10):3071–3082. [PubMed: 21975590]

Gracco VL, Löfqvist A. Speech Motor Coordination and Control: Evidence from Lip, Jaw, and Laryngeal Movements. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 1994; 14(11 Pt 1):6585–6597. [PubMed: 7965062]

Groothuis, Jitte; Ramsey, Nick F; Ramakers, Geert MJ; van der Plasse, Geoffrey. Physiological Challenges for Intracortical Electrodes. Brain Stimulation. 2014; 7(1):1–6. [PubMed: 23941984]

Guenther, Frank H, Brumberg, Jonathan S, Wright, E Joseph; Nieto-Castanon, Alfonso; Tourville, Jason A, Panko, Mikhail; Law, Robert; , et al. A Wireless Brain-Machine Interface for Real-Time Speech Synthesis. In: Ben-Jacob, Eshel, editor. PLoS ONE. Vol. 4. 2009. e8218

Halgren, Eric; Marinkovic, Ksenija; Chauvel, Patrick. Generators of the Late Cognitive Potentials in Auditory and Visual Oddball Tasks. Electroencephalography and Clinical Neurophysiology. 1998; 106(2):156–164. [PubMed: 9741777]

Herff, Christian; Heger, Dominic; de Pesters, Adriana; Telaar, Dominic; Brunner, Peter; Schalk, Gerwin; Schultz, Tanja. Brain-to-Text: Decoding Spoken Phrases from Phone Representations in the Brain. Frontiers in Neuroscience. 2015 Jun.

Hickok, Gregory; Poeppel, David. The Cortical Organization of Speech Processing. Nature Reviews Neuroscience. 2007; 8(5):393–402. [PubMed: 17431404]

Hochberg, Leigh R; Bacher, Daniel; Jarosiewicz, Beata; Masse, Nicolas Y; Simeral, John D; Vogel, Joern; Haddadin, Sami; , et al. Reach and Grasp by People with Tetraplegia Using a Neurally Controlled Robotic Arm. Nature. 2012; 485(7398):372–375. [PubMed: 22596161]

Houde, John F; Chang, Edward F. The Cortical Computations Underlying Feedback Control in Vocal Production. Current Opinion in Neurobiology. 2015 Aug.33:174–181. [PubMed: 25989242]

Huang, Jie; Carr, Thomas H; Cao, Yue. Comparing Cortical Activations for Silent and Overt Speech Using Event-Related fMRI. Human Brain Mapping. 2002; 15(1):39–53. [PubMed: 11747099]

Huth, Alexander G; Nishimoto, Shinji; Vu, An T; Gallant, Jack L. A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. Neuron. 2012; 76(6):1210–1224. [PubMed: 23259955]

Ikeda, Shigeyuki; Shibata, Tomohiro; Nakano, Naoki; Okada, Rieko; Tsuyuguchi, Naohiro; Ikeda, Kazushi; Kato, Amami. Neural Decoding of Single Vowels during Covert Articulation Using Electrocorticography. Frontiers in Human Neuroscience. 2014:125. [PubMed: 24639642]

Indefrey P, Levelt WJM. The Spatial and Temporal Signatures of Word Production Components. Cognition. 2004; 92(1–2):101–144. [PubMed: 15037128]

Kay, Kendrick N; Gallant, Jack L. I Can See What You See. Nature Neuroscience. 2009; 12(3):245. [PubMed: 19238184]

Kellis, Spencer; Miller, Kai; Thomson, Kyle; Brown, Richard; House, Paul; Greger, Bradley. Decoding Spoken Words Using Local Field Potentials Recorded from the Cortical Surface. Journal of Neural Engineering. 2010; 7(5):056007. [PubMed: 20811093]

King, Simon. An Introduction to Statistical Parametric Speech Synthesis. Sadhana. 2011; 36(5):837–852.

Kubanek, Jan; Brunner, Peter; Gunduz, Aysegul; Poeppel, David; Schalk, Gerwin. The Tracking of Speech Envelope in the Human Cortex. In: Rodriguez-Fornells, Antoni, editor. PLoS ONE. Vol. 8. 2013. e53398

Kubánek J, Miller KJ, Ojemann JG, Wolpaw JR, Schalk G. Decoding Flexion of Individual Fingers Using Electrocorticographic Signals in Humans. Journal of Neural Engineering. 2009; 6(6): 066001. [PubMed: 19794237]

Lachaux, Jean-Philippe; Axmacher, Nikolai; Mormann, Florian; Halgren, Eric; Crone, Nathan E. High-Frequency Neural Activity and Human Cognition: Past, Present and Possible Future of Intracranial EEG Research. Progress in Neurobiology. 2012; 98(3):279–301. [PubMed: 22750156]

Leonard MK, Bouchard KE, Tang C, Chang EF. Dynamic Encoding of Speech Sequence Probability in Human Temporal Cortex. Journal of Neuroscience. 2015; 35(18):7203–7214. [PubMed: 25948269]

Leuthardt, Eric C; Pei, Xiao-Mei; Breshears, Jonathan; Gaona, Charles; Sharma, Mohit; Freudenberg, Zac; Barbour, Dennis; Schalk, Gerwin. Temporal Evolution of Gamma Activity in Human Cortex during an Overt and Covert Word Repetition Task. Frontiers in Human Neuroscience. 2012; 6

Leuthardt, G Schalk; Wolpaw, JR; Ojemann, JG; Moran, DW. A Brain–computer Interface Using Electrocorticographic Signals in Humans. Journal of Neural Engineering. 2004; 1(2):63–71. [PubMed: 15876624]

Levelt. Speaking: From Intention to Articulation. Bradford Books, U.S.; 1993. https://books.google.com/books?id=LbVCdCE-NQAC

Levelt WJ, Roelofs A, Meyer AS. A Theory of Lexical Access in Speech Production. The Behavioral and Brain Sciences. 1999; 22(1):1–38. [PubMed: 11301520]

Luo, Huan; Poeppel, David. Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. Neuron. 2007; 54(6):1001–1010. [PubMed: 17582338]

Manning JR, Jacobs J, Fried I, Kahana MJ. Broadband Shifts in Local Field Potential Power Spectra Are Correlated with Single-Neuron Spiking in Humans. Journal of Neuroscience. 2009; 29(43): 13613–13620. [PubMed: 19864573]

Martin, Stephanie; Brunner, Peter; Holdgraf, Chris; Heinze, Hans-Jochen; Crone, Nathan E; Rieger, Jochem; Schalk, Gerwin; Knight, Robert T; Pasley, Brian N. Decoding Spectrotemporal Features of Overt and Covert Speech from the Human Cortex. Frontiers in Neuroengineering. 2014

MesgaraniChang, Edward F. Selective Cortical Representation of Attended Speaker in Multi-Talker Speech Perception. Nature. 2012; 485(7397):233–236. [PubMed: 22522927]

Mesgarani N, Cheung C, Johnson K, Chang EF. Phonetic Feature Encoding in Human Superior Temporal Gyrus. Science. 2014; 343(6174):1006–1010. [PubMed: 24482117]

Millán, Galan F, Vanhooydonck D, Lew E, Philips J, Nuttin M. Asynchronous Non-Invasive Brain-Actuated Control of an Intelligent Wheelchair. Conference Proceedings: … Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference. 2009; 2009:3361–3364.

Minev IR, Musienko P, Hirsch A, Barraud Q, Wenger N, Moraud EM, Gandar J, et al. Electronic Dura Mater for Long-Term Multimodal Neural Interfaces. Science. 2015; 347(6218):159–163. [PubMed: 25574019]

Mugler, Emily M; Patton, James L; Flint, Robert D; Wright, Zachary A; Schuele, Stephan U; Rosenow, Joshua; Shih, Jerry J; Krusienski, Dean J; Slutzky, Marc W. Direct Classification of All American English Phonemes Using Signals from Functional Speech Motor Cortex. Journal of Neural Engineering. 2014; 11(3):035015. [PubMed: 24836588]

Naselaris, Thomas; Kay, Kendrick N; Nishimoto, Shinji; Gallant, Jack L. Encoding and Decoding in fMRI. NeuroImage. 2011; 56(2):400–410. [PubMed: 20691790]

Nishimoto, Shinji; Vu, An T; Naselaris, Thomas; Benjamini, Yuval; Yu, Bin; Gallant, Jack L. Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Current Biology. 2011; 21(19):1641–1646. [PubMed: 21945275]

Palmer, Erica D; Rosen, Howard J; Ojemann, Jeffrey G; Buckner, Randy L; Kelley, William M; Petersen, Steven E. An Event-Related fMRI Study of Overt and Covert Word Stem Completion. NeuroImage. 2001; 14(1):182–193. [PubMed: 11525327]

Paninski, Liam; Pillow, Jonathan; Lewi, Jeremy. Progress in Brain Research. Vol. 165. Elsevier; 2007. Statistical Models for Neural Encoding, Decoding, and Optimal Stimulus Design; 493–507. http://linkinghub.elsevier.com/retrieve/pii/S0079612306650310

Parneet, Kaur; Garg, Vidushi. Speech Recognition System; Challenges and Techniques. International Joural of Computer Science and Infomration Technlogies. 2012; 3

Pasley, Brian N, David, Stephen V, Mesgarani, Nima; Flinker, Adeen; Shamma, Shihab A, Crone, Nathan E, Knight, Robert T, Chang, Edward F. Reconstructing Speech from Human Auditory Cortex. In: Zatorre, Robert, editor. PLoS Biology. Vol. 10. 2012. e1001251

Pasley, Brian N, Knight, Robert T. Progress in Brain Research. Vol. 207. Elsevier; 2013. Decoding Speech for Understanding and Treating Aphasia; 435–456. http://linkinghub.elsevier.com/retrieve/pii/B9780444633279000187

Pei, Xiaomei; Barbour, Dennis L; Leuthardt, Eric C; Schalk, Gerwin. Decoding Vowels and Consonants in Spoken and Imagined Words Using Electrocorticographic Signals in Humans. Journal of Neural Engineering. 2011 Aug.(4)

Perrone-Bertolotti M, Rapin L, Lachaux J-P, Baciu M, Loevenbruck H. What Is That Little Voice inside My Head? Inner Speech Phenomenology, Its Role in Cognitive Performance, and Its Relation to Self-Monitoring. Behavioural Brain Research. 2014 Mar.261:220–239. [PubMed: 24412278]

Price. The Anatomy of Language: Contributions from Functional Neuroimaging. Journal of Anatomy. 2000; 197(3):335–359. [PubMed: 11117622]

Price CJ. A Review and Synthesis of the First 20 years of PET and fMRI Studies of Heard Speech, Spoken Language and Reading. NeuroImage. 2012; 62(2):816–847. [PubMed: 22584224]

Rapp B, Fischer-Baum S, Miozzo M. Modality and Morphology: What We Write May Not Be What We Say. Psychological Science. 2015; 26(6):892–902. [PubMed: 25926478]

Rissman J, Greely HT, Wagner AD. Detecting Individual Memories through the Neural Decoding of Memory States and Past Experience. Proceedings of the National Academy of Sciences. 2010; 107(21):9849–9854.

Robson, Holly; Grube, Manon; Lambon Ralph, Matthew A; Griffiths, Timothy D; Sage, Karen. Fundamental Deficits of Auditory Perception in Wernicke's Aphasia. Cortex. 2013; 49(7):1808–1822. [PubMed: 23351849]

Rouse AG, Williams JJ, Wheeler JJ, Moran DW. Cortical Adaptation to a Chronic Micro-Electrocorticographic Brain Computer Interface. Journal of Neuroscience. 2013; 33(4):1326–1330. [PubMed: 23345208]

Schalk G, Kubánek J, Miller KJ, Anderson NR, Leuthardt EC, Ojemann JG, Limbrick D, Moran D, Gerhardt LA, Wolpaw JR. Decoding Two-Dimensional Movement Trajectories Using Electrocorticographic Signals in Humans. Journal of Neural Engineering. 2007; 4(3):264–275. [PubMed: 17873429]

Shamma, Shihab. Physiological Foundations of Temporal Integration in the Perception of Speech. Journal of Phonetics. 2003; 31(3–4):495–501.

Shuster, Linda I; Lemieux, Susan K. An fMRI Investigation of Covertly and Overtly Produced Mono- and Multisyllabic Words. Brain and Language. 2005; 93(1):20–31. [PubMed: 15766765]

Singleton, Nina Capone; Shulman, Brian B. Language Development: Foundations, Processes, and Clinical Applications. 2nd. Burlington, MA: Jones & Bartlett Learning; 2014.

Slutzky, Marc W; Jordan, Luke R; Krieg, Todd; Chen, Ming; Mogul, David J; Miller, Lee E. Optimal Spacing of Surface Electrode Arrays for Brain–machine Interface Applications. Journal of Neural Engineering. 2010; 7(2):026004.

Staba, Richard J; Wilson, Charles L; Bragin, Anatol; Fried, Itzhak; Engel, Jerome. Quantitative Analysis of High-Frequency Oscillations (80–500 Hz) Recorded in Human Epileptic Hippocampus and Entorhinal Cortex. Journal of Neurophysiology. 2002; 88(4):1743–1752. [PubMed: 12364503]

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: Computation in Neural Systems. 2001; 12(3):289–316.

Tomasello, Michael. Origins of Human Communication. Cambridge, Mass: MIT Press; 2008. The Jean Nicod Lectures 2008

Trask, RL. Language: The Basics. 2nd. London ; New York: Routledge; 1999.

Varoquaux, Gael; Thirion, Bertrand. How Machine Learning Is Shaping Cognitive Neuroimaging. GigaScience. 2014; 3(1):28. [PubMed: 25405022]

Vitevitch, Michael S; Luce, Paul A; Pisoni, David B; Auer, Edward T. Phonotactics, Neighborhood Activation, and Lexical Access for Spoken Words. Brain and Language. 1999; 68(1–2):306–311. [PubMed: 10433774]

Vu, Vincent Q; Ravikumar, Pradeep; Naselaris, Thomas; Kay, Kendrick N; Gallant, Jack L; Yu, Bin. Encoding and Decoding V1 fMRI Responses to Natural Images with Sparse Nonparametric Models. The Annals of Applied Statistics. 2011; 5(2B):1159–1182. [PubMed: 22523529]

Waibel, Alex; Lee, Kai-Fu, editors. Readings in Speech Recognition. San Mateo, Calif: Morgan Kaufmann Publishers; 1990.

Watkins KE, Dronkers NF, Vargha-Khadem F. Behavioural Analysis of an Inherited Speech and Language Disorder: Comparison with Acquired Aphasia. Brain: A Journal of Neurology. 2002; 125(Pt 3):452–464. [PubMed: 11872604]

Wei WangDegenhart, AD, Sudre, GP, Pomerleau, DA, Tyler-Kabara, EC. Decoding Semantic Information from Human Electrocorticographic (ECoG) Signals. IEEE; 2011. 6294–6298.

Winkler, István; Denham, Susan L; Nelken, Israel. Modeling the Auditory Scene: Predictive Regularity Representations and Perceptual Objects. Trends in Cognitive Sciences. 2009; 13(12):532–540. [PubMed: 19828357]

Wodlinger, B, Degenhart, AD, Collinger, JL, Tyler-Kabara, EC, Wei Wang. The Impact of Electrode Characteristics on Electrocorticography (ECoG). IEEE; 2011. 3083–3086.

Wolpaw, Jonathan R; Birbaumer, Niels; McFarland, Dennis J; Pfurtscheller, Gert; Vaughan, Theresa M. Brain-Computer Interfaces for Communication and Control. Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology. 2002; 113(6):767–791. [PubMed: 12048038]

Wolpaw, McFarland DJ, Neat GW, Forneris CA. An EEG-Based Brain-Computer Interface for Cursor Control. Electroencephalography and Clinical Neurophysiology. 1991; 78(3):252–259. [PubMed: 1707798]

Wu, Michael C-K; David, Stephen V; Gallant, Jack L. Complete Functional Characterization of Sensory Neurons by System Identification. Annual Review of Neuroscience. 2006; 29:477–505.

Zion, Golumbic; Nai Ding, Elana M; Bickel, Stephan; Lakatos, Peter; Schevon, Catherine A; McKhann, Guy M; Goodman, Robert R; , et al. Mechanisms Underlying Selective Neuronal

Tracking of Attended Speech at a ‘Cocktail Party.’. Neuron. 2013; 77(5):980–991. [PubMed: 23473326]

Highlights

- Functional organization of language.
- Properties of intracranial recordings.
- Approaches to decoding features of speech perception and production.
- Decoding of imagined speech.
- Challenges and opportunities in decoding human language features.

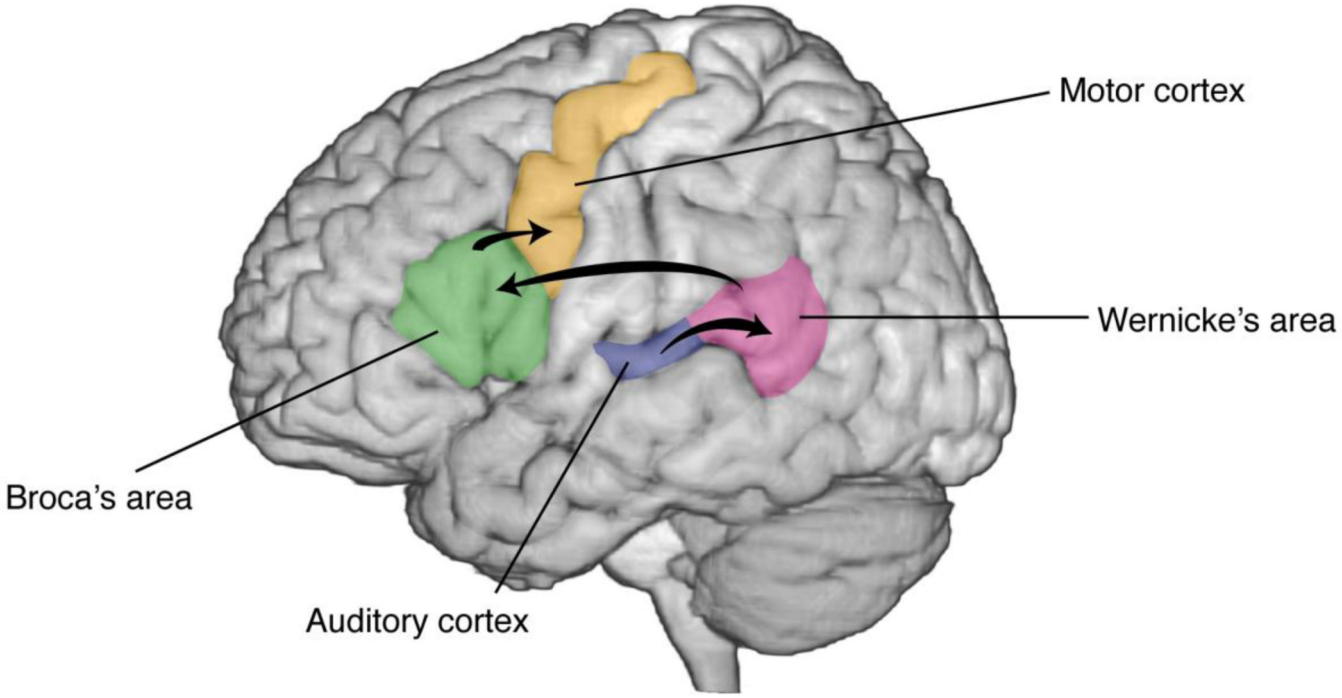

Figure 1.

Speech network. The major brain areas involved in speech processing are depicted (Price 2000). Early auditory cortices (Heschl's gyrus) involve spectrotemporal analysis of speech, and project into Wernicke's area (posterior superior temporal gyrus). The arcuate fasciculus connects Wernicke's area to Broca's area, which is involved in speech preparation and planning. Finally, the ventral sensorimotor cortex coordinates articulatory movements to produce audible speech.

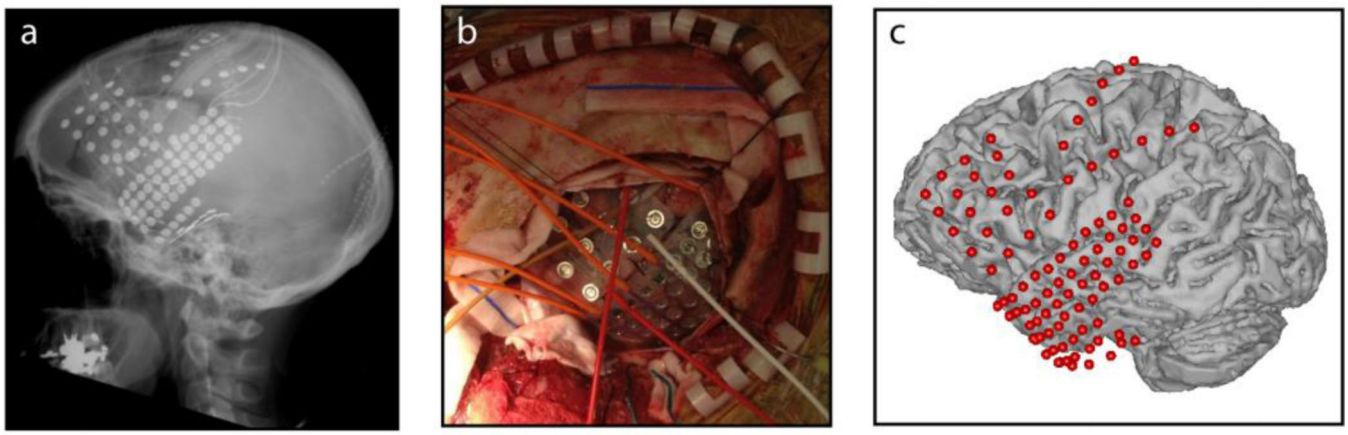

Figure 2.
The ECoG grid and surgical placement (a) ECoG surgical placement. (b) Radiography of electrode placement. (c) Electrode positions in situ.

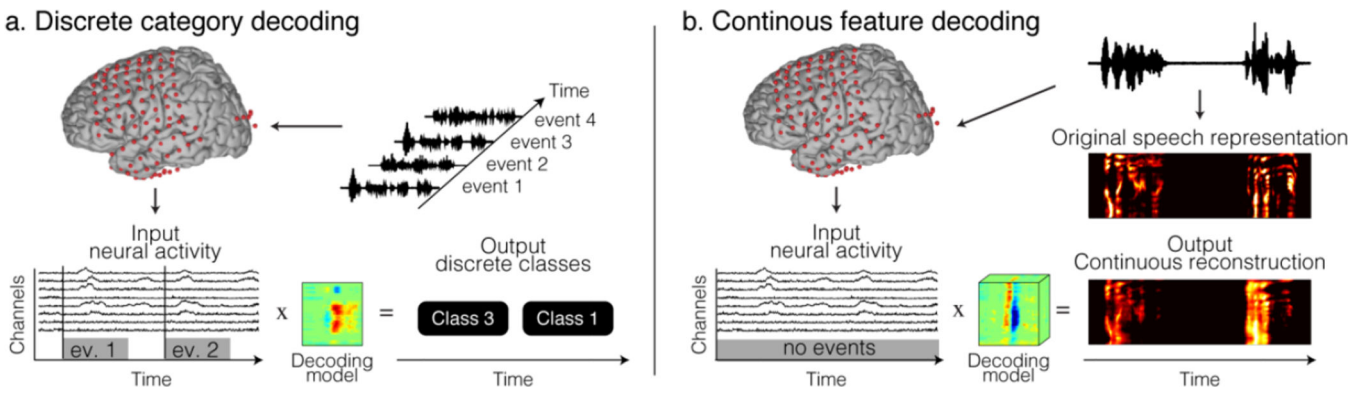

Figure 3.

Neural decoding a) In discrete category decoding, events in the neural activity are classified into a finite number of choices. b) In continuous feature decoding, features of speech are modeled and reconstructed continuously over time.

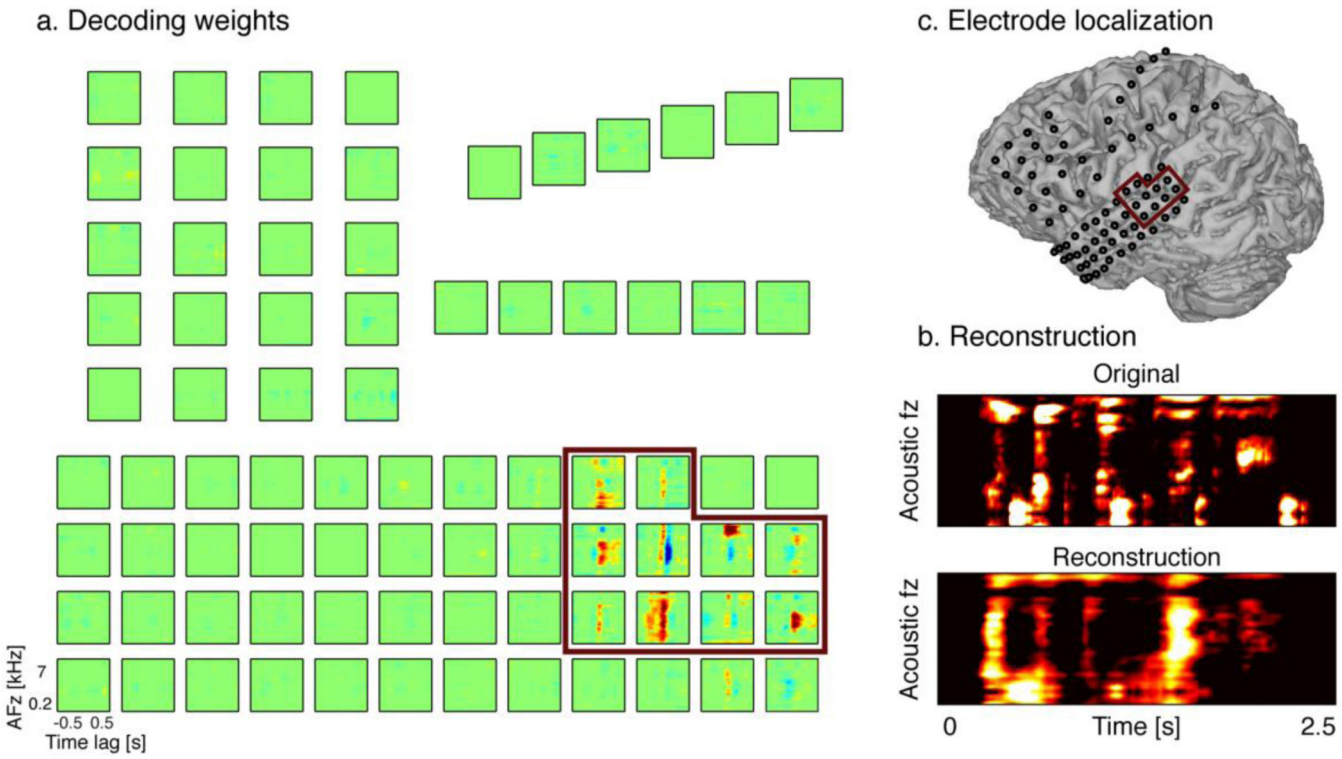

Figure 4.

Reconstruction a) Decoding weights for each electrodes b) Electrode location c) Example of original and reconstructed spectrograms.

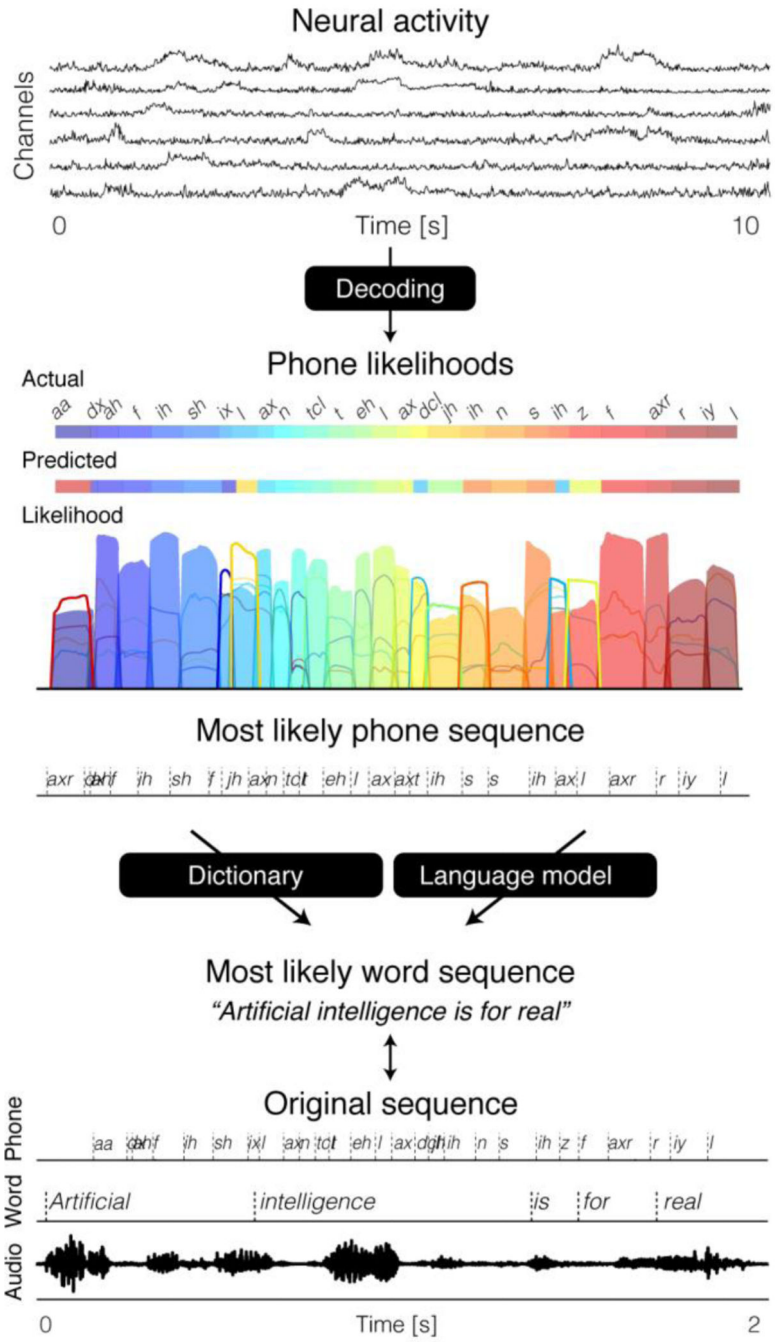

Figure 5.

Phonetic decoding procedure. Neural activity continuously decoded into phone likelihoods, and subsequently translated into most likely discrete phone sequence. During the decoding procedure, the information about the observed neural activity was combined with statistical linguistic information – in order to find the most likely sequence of phones and words, given the observed neural activity sequence. The decoded sequence is then compared with the original sequence.

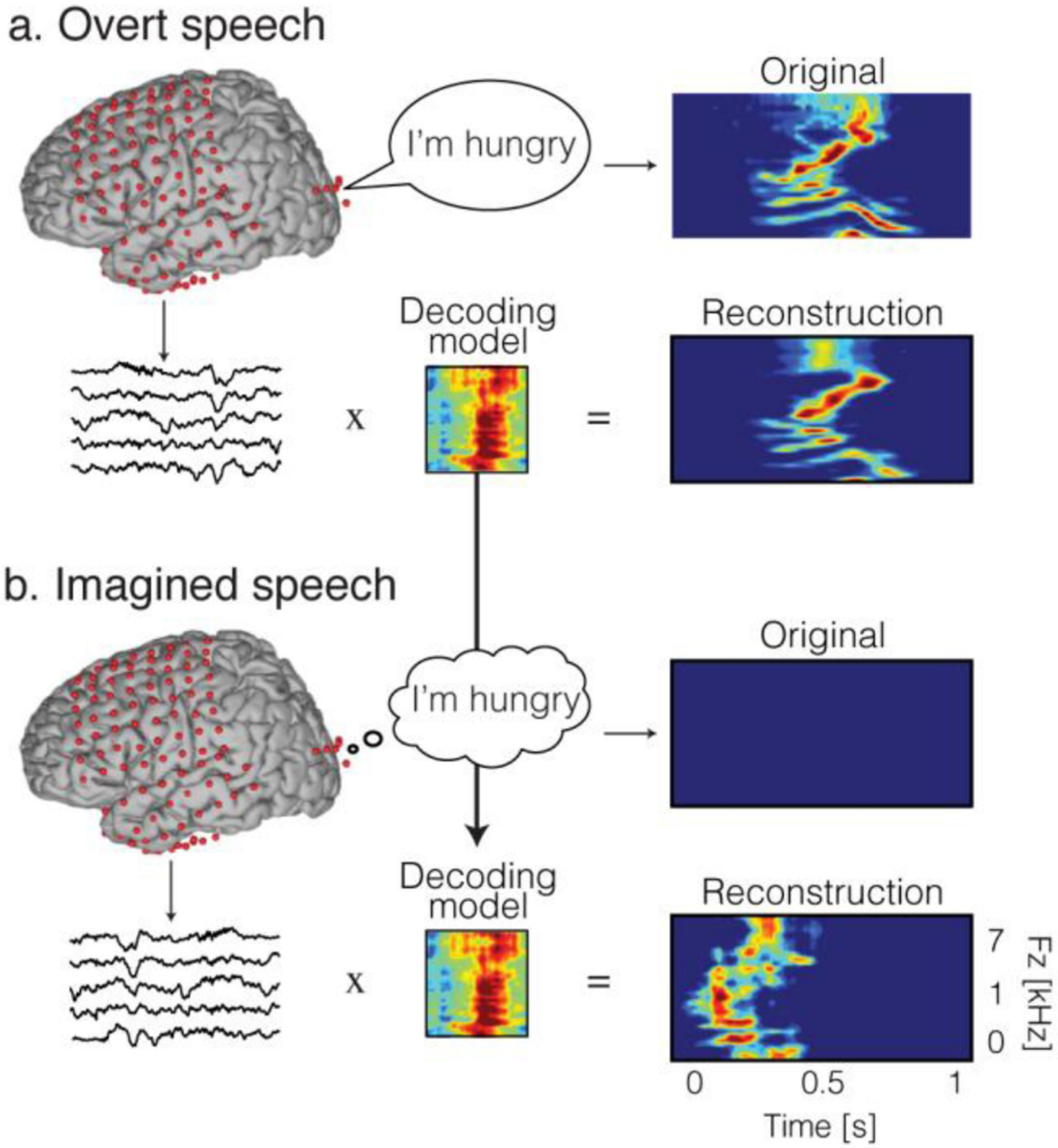

Figure 6.

Imagined speech decoding approach. a) The decoding model is built using data generated during an overt speech production task. b) The decoding model built in a) is applied to neural data in the imagined speech condition. The reconstructed spectrogram in the imagined speech is compared to the corresponding original overt speech spectrogram.