Biomed Eng Lett (2015) 5:10-21
DOI 10.1007/s13534-015-0175-1

REVIEW ARTICLE

Progress in Speech Decoding from the Electrocorticogram

Shreya Chakrabarti, Hilary M. Sandberg, Jonathan S. Brumberg and Dean J. Krusienski

Received: 27 February 2015 / Revised: 7 March 2015 / Accepted: 9 March 2015

Abstract
Recent advances in neuroimaging methods have improved our ability to explore the neurological processes underlying speech and language. As a result of these investigations, it is now possible to decode aspects of speech directly from neural activity toward the development of neuroprosthetic devices for individuals with severe neuromuscular and communication disorders. Much of what is known about the neural correlates of speech articulation and perception is based on lesion and cortical electrical stimulation studies, as well as modern non-invasive neuroimaging. Though extremely important to the current understanding of brain function, traditional neuroimaging methods are primarily limited by the spatial and temporal resolution of the imaging technique. Electrical activity measured from the cortex, or electrocorticography (ECoG), offers several advantages over other neuroimaging modalities for characterization and real-time decoding of brain activity. Specifically, ECoG is well-suited for the study of speech and language owing to its unique spatial and temporal resolution capabilities that allow it to accurately capture the fast-changing dynamics of the large cortical networks underlying speech processing. This review presents the current progress of ECoG-based speech characterization and decoding studies, including an overview of prior neuroimaging studies, ECoG representations of speech production and perception, and a discussion of future directions.

Keywords Electrocorticography, ECoG, Speech, Neuroprosthetics

Shreya Chakrabarti (✉), Dean J. Krusienski
Dept of Electrical and Computer Engineering, Old Dominion University, Norfolk, VA, USA
Tel : +17576833752 / Fax : +17576833220
E-mail : schak001@odu.edu

Hilary M. Sandberg
Dept of Communication Sciences and Disorders, Old Dominion University, Norfolk, VA, USA

Jonathan S. Brumberg
Dept of Speech-Language-Hearing: Sciences & Disorders, University of Kansas, Lawrence, KS, USA

INTRODUCTION

While the neural correlates of speech processing have been investigated for several decades, recent research has focused on investigating the possibility of decoding speech directly from neural activity. Currently, close to two million people in the United States, and far more worldwide, have significant verbal communication deficits as a result of severe neuromuscular impairments due to injury or disease [1, 2]. Communication impairments can originate from neurodegenerative disorders that affect the motor production and articulation of speech, such as amyotrophic lateral sclerosis (ALS), or from language disorders that affect the cognitive production or comprehension of language, such as various forms of aphasia [3]. One goal of characterizing neural activity during speech production and comprehension is to develop neurotechnological applications to restore communication to those affected by speech and language disorders. The majority of neuroimaging studies of communication have used functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) to localize the neuroanatomy involved in speech and language [4], while noninvasive electromagnetic techniques such as magnetoencephalography (MEG) and electroencephalography (EEG) have additionally provided information related to the temporal patterns of activation. In particular, fMRI and PET have greatly contributed to the understanding of speech processing in the human brain. However, such techniques that rely on measurements of hemodynamic responses (timescale of 4−6 seconds) are unable to capture the rapid temporal dynamics of natural speech (phoneme productions are often < 200 ms). Electrocorticography (ECoG), the measurement of electrical activity from the cortex, has become a highly promising neural signal acquisition modality for studying speech and language processing due to its capability to provide high spatial and temporal resolution [5]. Ultimately, the unique information offered by ECoG can potentially be used to develop neuroprostheses that will decode intended speech for expressive communication or represent perceived language information directly using neural activity to augment receptive communication. This review begins with an overview of prior neuroimaging and lesion studies that serve as foundation for the neural correlates of language. ECoG is then introduced with an emphasis on recent speech characterization and decoding studies. Finally, the review concludes with a discussion of future directions of speech research and development using ECoG.

Revealing neural correlates of language

Prior to the development of functional neuroimaging techniques, identification of language areas in the human brain was based on studies of deficits in patients with neurological damage or patients undergoing electrical stimulation during neurosurgery [6]. The process of identifying parts of the brain involved in language processing began as early as 1861, when neurosurgeon Paul Broca studied the brains of nine patients with lesions and concluded that the expressive language centers present in most humans are located in the posterior inferior frontal gyrus of the left hemisphere, in an area now known as Broca's area [7]. A decade later, another renowned neurologist, Carl Wernicke, discovered that the posterior part of the left temporal lobe is involved in the comprehension of language. This region is now known as Wernicke's area [8]. Other researchers have also developed functional models of speech that describe the speech-related neural areas and their functional significance [9, 10]. These models have identified the functional network consisting of the pre-motor cortex, primary motor cortex, Broca's area, primary auditory cortex, Wernicke's area, and superior temporal gyrus (STG) to be involved in the planning and production of speech and in the perception of speech (see Fig. 1).

Neuroimaging techniques such as PET or fMRI are also used to identify the neural correlates of speech processing by the human brain [4, 11-15]. PET imaging involves the use of positron-emitting isotopes as tracers to detect changes in cerebral blood flow and volume in response to a stimulus [16]. For speech processing, a stimulus might be the presentation of acoustic speech or the preparation and execution of a speech motor task. The spatial resolution in PET imaging is approximately 6 mm, while the temporal resolution is between tens of seconds to several minutes. In contrast, fMRI imaging results from changes in the blood oxygenation level due to the metabolic activity of neuronal tissue without the use of a radioactive tracer. The spatial resolution of fMRI ranges from 1–4 mm and the temporal resolution is between hundreds of milliseconds to seconds.

Studies have investigated the neural bases of speech production as well as the perception of speech using scalp EEG [17-19]. Due to the comparatively low signal-to-noise ratio of EEG, time-locked averages known as event-related potentials (ERPs) are needed to capture the relevant brain responses. ERPs have been successfully used to study the temporal evolution of the phonological and lexical processes in the human brain corresponding to speech production and perception [17-19], but not during active speech production due to contamination from myoelectrical artifacts. While noninvasive measures such as EEG and MEG can theoretically offer adequate temporal resolution, factors such as spatial resolution on the order of centimeters, spectral bandwidth on the order of 80 Hz, and susceptibility to electrical artifacts severely limit the utility of these modalities for investigating the joint spatiotemporal dynamics of brain activity.

The excellent spatial resolution of fMRI and PET and the temporal resolution of EEG and MEG have provided theoretical and computational models that highlight the spatial topography and functional connectivity of the brain networks involved in speech production and comprehension [4, 17-20]. The results of these studies highlight the interconnectedness of neural networks for speaking, of which the traditional Broca's and Wernicke's areas play a crucial role in production and perception, respectively [4, 20]. New advances in intracranial measurements of electrical activity such as ECoG can simultaneously capture the spatial and temporal neurological dynamics of speech and language processing, with respective spatial and temporal resolutions that are equivalent or superior to those obtained using standard noninvasive neuroimaging and electrophysiology techniques. This provides a new opportunity to observe and characterize the cortical speech network during production and perception tasks, which can lead toward the development of improved real-time decoding models and eventually neural prostheses for speech and language applications.

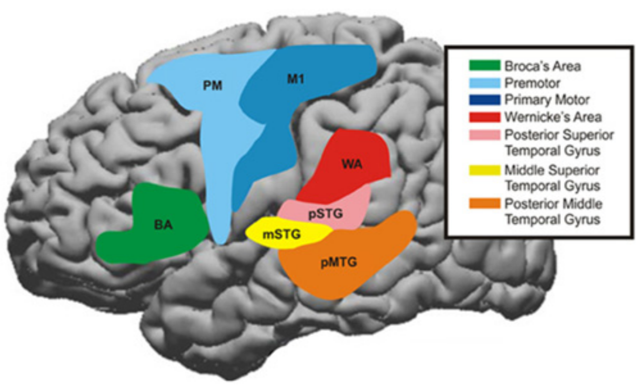

Fig. 1. The areas of the cortex involved in speech perception and speech articulation, shown on a generic brain using color maps that show approximate locations of the different areas. The pre-motor area, primary motor area, and Broca's area are involved in speech preparation and articulation while the posterior and middle superior temporal gyrus, the posterior middle temporal gyrus and Wernicke's area are associated with speech perception and processing. Adapted with permission from reference [21].

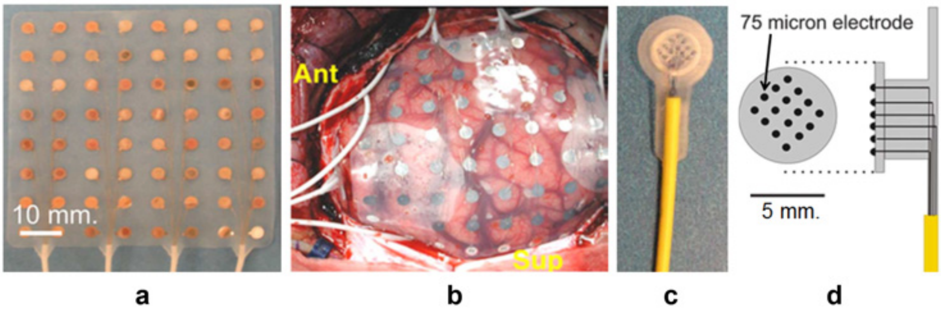

Fig. 2. Macro and micro ECoG arrays. (a) Standard clinical macrogrid. (b) Surgical placement of macrogrid. (c) Microgrid array. (d) Schematic of microgrid array. Adapted with permission from reference [23].

The electrocorticogram

Signal acquisition and characteristics

ECoG measures the electrical activity of the brain recorded by electrodes placed directly on the surface of the cortex. ECoG was originally developed by neurosurgeons W. Penfield and H. Jasper in the 1930's as a technique for localization of epileptic seizure foci prior to surgical resection [22]. Modern ECoG typically uses platinum electrodes with a diameter of 4 mm that are implanted as either two-dimensional grids (e.g., 8×8 electrodes) or one-dimensional arrays (e.g., 4 or 6 electrodes) with an inter-electrode distance of 10 mm [23], as shown in Fig. 2. In addition to standard clinical ECoG arrays, micro-ECoG arrays (center-to-center distance of 4 mm or less) have also been used in recent studies to improve spatial resolution [24-27]. An example of a microgrid array is also shown in Fig. 2.

ECoG recordings are well-suited for basic neuroscience research as well as for neural decoding studies. The recording characteristics of ECoG include: (1) spatial resolution on the scale of millimeters (activity is related to the neural tissue directly beneath the electrode disk), (2) frequency bandwidth up to 200 Hz or higher, (3) an amplitude up to 100 μV compared to near 20 μV for EEG, and (4) reduced sensitivity to movement and myoelectrical artifacts compared to EEG and MEG [23]. Due to higher signal amplitudes and lower sensitivity to artifacts, ECoG signals also have higher signal-to-noise ratios (SNRs) compared to EEG. Compared to electrodes that penetrate the cortex, it is believed that ECoG is less likely to suffer from adverse tissue reactions and electrode encapsulation issues [28] that can degrade signal quality over time because ECoG does not breach the cortex [29, 30]. The superior bandwidth of ECoG compared to EEG is particularly important because modulations of the 70–180 Hz gamma-band range have been shown to be highly correlated with perception, cognitive function, and motor tasks [31-35]; including learning, memory and speech [21, 36-57]. These studies demonstrate the promising real-time decoding potential of ECoG.

Signal processing and feature extraction

To date, the most relevant information or features of ECoG are based on its spectral dynamics. ECoG recordings typically require preprocessing to condition the signals for further analysis. A spatial common average reference (CAR) filter is commonly applied to remove any undesirable noise, fluctuations, and artifacts that may be present in all channels over a region [36, 37, 43, 46]. Signals are then high-pass filtered starting between 0.5–2 Hz to further attenuate low-frequency fluctuations and heartbeat artifacts [44, 46]. The signals are also notch or comb filtered at harmonics of 60 Hz (or 50 Hz as appropriate) to eliminate power line interference [40, 42-46]. Furthermore, any trials or channels that show excessive fluctuations, presence of outliers, low SNR, or are overlying pathological tissue are generally removed to ensure that these trials or channels do not bias the analysis [40, 44, 46].

Following the preprocessing stage, more advanced signal processing techniques are used to extract the relevant spectro-temporal features for further analysis. The signals are typically transformed to the frequency domain using a Discrete Fourier Transform (DFT), auto-regressive (AR) model, or band-power filtering [36, 37, 43]. To date, the gamma band, from approximately 30 Hz to 200 Hz, has provided the most informative description of the neural processes underlying speech. Within this frequency range, modulations of the high-gamma band, from approximately 60–80 Hz to 150–200 Hz, have been identified as highly correlated with speech production and perception. This range of frequencies is of particular interest in ECoG because it is not detectable with the limited spectral ranges of other neural signal acquisition techniques such as scalp EEG [23]. Nevertheless, ECoG is also well suited to examine the delta (<4 Hz), theta (4–8 Hz), alpha (8–12 Hz) and beta bands (12–30 Hz) in the context of speech production and perception in the human cortex [21, 36], and often results in superior signal quality due to the reduction of electrical artifacts compared to EEG.

The extracted spectro-temporal ECoG features have been used to both study and decode neural activity during speech production and perception. For studies of the spatial characteristics of the speech network, recorded signals from electrodes at different locations over the cortex are compared against a reference signal (e.g., recorded speech) using statistical techniques such as correlation analysis or analysis of variance (ANOVA) [21, 36, 37, 40, 43, 45-48]. To obtain the temporal dynamics of these spatial networks, ECoG signals from each channel are compared to the reference signal at different time latencies relative to the onset of speech articulation or the presentation of speech stimuli to quantify the neural processing involved in the production and perception of speech, respectively [43, 45]. Decoding articulated or perceived speech from ECoG generally requires additional advanced signal processing and machine learning techniques to produce the desired speech outcome. Some recent attempts at speech decoding using these approaches are discussed in the 'ECoG-based decoding' section of this article.

Neural dynamics of speech and language processing using the electrocorticogram

Numerous studies have investigated cortical activity using ECoG during various speech tasks to identify the cortical areas and networks involved in speech production and perception, which create the groundwork for speech decoding. The following sections summarize the studies focusing on the identification of spatial and temporal dynamics of ECoG during speech production and perception.

Spatial characterization of speech production

A recent study by Bouchard *et al.* (2013) examined modulations of the ECoG high gamma-band during production of consonant-vowel syllables to investigate the phonetic organization of the speech sensorimotor cortex [41]. Spatial patterns of cortical activity showed that the gamma band activity recorded by electrodes over the sensorimotor cortex was present with different spatial organizations for consonants versus vowels. The spatial patterns also confirmed prior neuroimaging findings that speech is produced through the coordination of a distinct set of articulatory representations in the ventral sensorimotor cortex. In another study, Pei *et al.* (2011) analyzed the high-gamma power of ECoG signals recorded while subjects overtly or covertly repeated words presented acoustically or visually [37]. Overt word production was associated with high-gamma power changes in the superior and middle parts of the temporal lobe, Wernicke's area, Broca's area, the pre-motor cortex and the primary motor cortex. Covert word production, in contrast, was associated with high-gamma changes in the superior temporal lobe and the supramarginal gyrus. This study provided corroborating evidence for an overt speech production network identified in prior neuroimaging studies [4], but conflicted with prior accounts of covert speech, which merely suggested that the speech network was reduced in size and strength for the covert condition as compared to the overt condition [4, 12]. This study also demonstrated weaker and less distributed cortical activations during the covert speech condition, but in certain areas, in particular, the superior temporal lobe, no significant difference in activation was found between the overt and covert conditions. This study, thus, highlights the important role played by the superior temporal lobe during covert speech production. The neural correlates of verb generation and noun reading have also been investigated using an analysis of the ECoG high-gamma band [42]. The results of this study showed that activation was found in the primary mouth motor area, STG, and Broca's and Wernicke's areas, which agrees with previously identified regions involved in speech production.

Spatial characterization of speech perception

ECoG has also been used to explore the differences between the processing of speech and non-speech auditory stimuli, such as tones, in order to primarily highlight the importance of the human cortex in processing both the acoustic and the phonological aspects of speech. One of the early ECoG speech studies based on speech perception by Crone *et al.* (2001) explored the temporal and spatial activations of the cortex in response to perception of speech (phonemes) and non-speech (tones) stimuli [36]. This study found that activations in the primary auditory cortex and STG occur in the gamma band, and to a higher extent in the high gamma band, during phoneme discrimination. For the tone stimuli, it was found that increases of the gamma power occurred in fewer electrodes (i.e., smaller spatial extent) and with lower magnitudes than for phoneme stimuli. This effect was particularly noticeable in the left STG, which highlights the importance of the left hemisphere auditory cortex in speech processing as compared to non-speech auditory processing, and confirms results from earlier lesion studies that investigated tone perception.

Additional studies have supported the claim that processing of incoming auditory stimuli results in an increase in high-gamma activations in the left STG [44], and have elucidated the importance of the speech envelope in speech comprehension [43, 45]. The speech envelope is the rectified speech waveform that fluctuates with speech intensity (e.g., loudness), phonetic content, and rhythmic cadence that is vital for understanding fluent, conversational speech. ECoG has the requisite spatial and temporal resolution to study this quickly changing speech signal, and prior work has found that ECoG in the belt areas of the auditory cortex (i.e. the areas lying relatively early in the auditory pathway) tracks modulations in the speech envelope as well [43]. This provides evidence that cortical signals closely represent the acoustic features of speech, and paves the way for future studies to investigate finer temporal aspects of speech processing in the cortex. An important study by Canolty *et al*. (2007) showed that the high-gamma activity in the ECoG tracked the spatiotemporal dynamics of word processing while subjects listened to a stream of verbs associated with the hand and the mouth [46]. From this study, it was found that the perception of verbs activates the posterior and middle STG as well as the superior temporal sulcus (STS), which supports previous studies that found evidence that the STG is largely involved in speech comprehension.

Other studies have attempted to investigate the cortical responses to altered speech feedback to identify the neural dynamics of sensory processing for error detection and correction. One particular study by Chang *et al*. (2013) recorded speech from subjects, and then later played back these recordings with slight perturbations in the pitch to the subjects while they were speaking [47]. It was found that the cortical responses in the posterior STG were suppressed while listening to unaltered feedback, but enhanced in response to the pitch-altered feedback, which corroborates results from a previous study which demonstrated the same effect in the EEG auditory response [58]. With ECoG it was possible to localize this change directly to the auditory cortex, while EEG only provided indirect evidence of auditory cortex involvement. The subjects were found to compensate for the altered pitch in the stimuli by changing the pitch of their speech. Furthermore, these vocal changes made by the subjects were predicted by their auditory cortical responses to the altered pitch stimuli. This neurological relationship provides evidence for the sensorimotor control of articulation in humans through the coordination of various cortical areas.

Temporal evolution of speech: from planning and production to perception

Recent studies have attempted to use language tasks that involve both speech production and perception to simultaneously analyze both expressive and receptive speech areas [48-50]. Analyzing high-gamma band ECoG activity, it was shown that the activation associated with listening was limited to STG and areas in and around Wernicke's area. Activation associated with spoken word production was found in the sensory or mouth motor regions as well as Broca's area. It can be concluded from the combination of all these studies that, broadly, different areas of the cortex are activated by motor speech production and speech perception, as illustrated in Fig. 1. The areas involved in speech articulation are the pre-motor cortex, which is mainly involved in planning; the face-mouth-motor regions, involved in generating mouth movements necessary for speech articulation; and Broca's area which is involved in speech planning and articulation. The areas of the cortex primarily involved in speech perception and comprehension are STG and Wernicke's area. Analysis of the temporal dynamics of ECoG signals during speech perception shows that the posterior STG is activated first, followed by the middle STG, and then the STS [46]. The spatial characterization results found by these studies analyzing both speech production and perception simultaneously further support results from the studies investigating speech production and speech perception independently, discussed in the previous two sections of this article, respectively.

ECoG-BASED DECODING

The aforementioned studies provide a framework for associating ECoG recordings (namely the high-gamma band power) with the behavioral tasks of speech production and perception. Fig. 3 illustrates the concept of training a neural-based decoding model using overt speech for reconstruction of imagined speech. This concept represents the basis of a speech neuroprosthetic device. The decoding model may

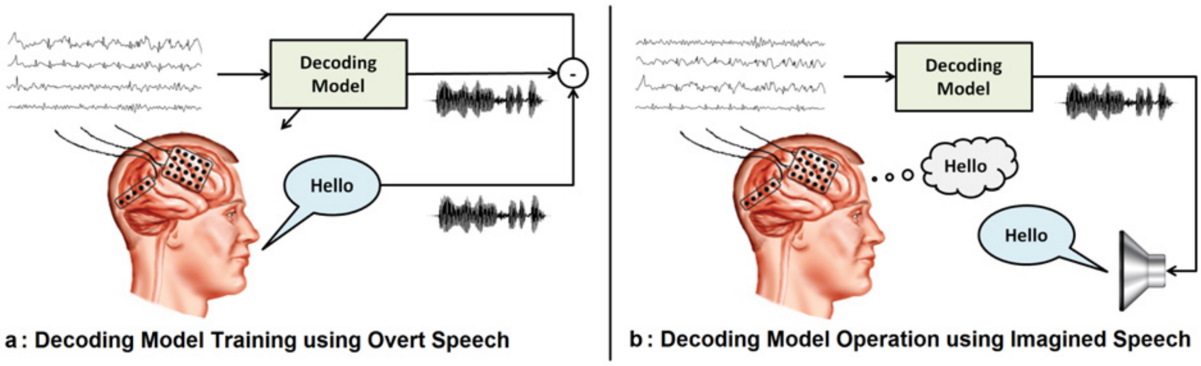

Fig. 3. Speech decoding model. (a) ECoG signals serve as the input to a neural-based decoding model that is trained using representations of recorded overt speech (e.g., time series, spectrogram, etc.) to ideally reconstruct the overt speech directly from the ECoG signals. (b) In principle, a version of the trained decoding model would be used to generate imagined speech directly from ECoG signals in real time. The model may be trained using overt speech or through some prior characterization of imagined speech, since there is no behavioral output during imagined speech.

perform a continuous reconstruction of the speech or a discrete classification and output of phonemes, words, etc., depending on the objective and constraints of the system. While it may be possible to decode individual words or phrases discretely, extending such models becomes highly dependent on the desired vocabulary and can become intractable. Alternatively, the ability to decode formants or phonemes will enable the creation of generative models that are not limited to a fixed vocabulary. In any case, effectively developing and transferring an overt speech-trained model to imagined speech remains an active research challenge since the neural representations of overt and imagined speech are not identical.

In terms of decoding perceptual information, one potential application lies in the development of a practical auditory neuroprosthesis. A practical auditory decoder should demonstrate the ability to segregate and process attended sound streams from irrelevant signals from more than one point in space in a complex acoustic environment, such as in a hospital setting or a restaurant. This ability would expand the active space of the signal, allowing for more natural communication when competing multi-talker streams exist in the acoustic space. Additionally, perceptual decoding could be used to enhance the speech production decoder by accounting for the production-perception feedback loop [47, 49]. The following sections discuss attempts to decode, or predict speech behavior directly from ECoG activity. These studies represent the initial steps toward the development of real-time neuroprostheses for speech and language.

Decoding of overt speech production

Speech production is a complex process that is initiated by linguistic processing and results in articulated speech, which can be further broken down into sentences, words, syllables, vowels, consonants and finally, into phonemes. The following speech decoding studies have investigated the decoding of these various levels of speech production. A study by Wang *et al*. (2011) showed an increase in gamma power over Broca's and Wernicke's area during picture naming and property identification [50] that was used to decode the semantic category associated with each stimulus. In this study, ECoG was recorded from four subjects as they participated in a picture naming task, where they were presented with pictures of objects from different semantic categories, such as food, tools, dwelling and body parts, varying from subject to subject. Two popularly used machine learning techniques, the Gaussian Naive Bayes Classifier and the linear support vector machine (SVM), were then used to decode the semantic category that the subjects named from the ECoG activity, resulting in accuracies as high as 74% (chance level 33%).

A study by Kellis *et al*. (2010), which involved the use of micro-ECoG electrodes implanted over the facial motor cortex and Wernicke's area, attempted to classify 45 possible pairs among a set of 10 words that a subject was articulating. The power spectra of the ECoG data and principal component analysis (PCA) were used to maximize the variance between the classes of words being identified [24]. The electrodes that led to the best classifier performance were selected to improve classification of word pairs, which demonstrated that electrodes implanted over the face-motor cortex resulted in better classification (40 of the 45 word pairs classified with an accuracy of 80% or higher, chance level 50%) than did the electrodes implanted over Wernicke's area (15 of the 45 word pairs classified with an accuracy of 80% or higher). This may be explained by the role of the face-motor cortex in the control of mouth movements required for speech articulation.

A study by Pei *et al*. (2011) examined ECoG signals to classify four vowels and nine pairs of consonants from a closed set of spoken whole words, and achieved 10-fold cross-validation classification accuracies up to 43% for vowels and 49% for consonant pairings, which were both significantly better than chance (chance level 25% for both) [51]. This study used the technique of maximum relevance and minimum redundancy (MRMR) to select the top 35 or 40 ECoG features for decoding consonant pairs or vowels respectively, using the Naive Bayes Classifier. A recent study by Kanas *et al*. (2014) performed spatio-spectral feature clustering of ECoG recordings in order to detect speech activity from one subject during a syllable repetition task, achieving an accuracy of 98.8% [52]. This study used the method of k-means clustering to group the best power spectral density features for all the channels and frequencies into clusters. This grouping was followed by the classification of the ECoG features using different algorithms, among which the support vector machine was the most accurate in detecting speech activity from the related cortical signals. Although the decoding performed by this study was limited to only speech versus non-speech detection, it highlights the cortical areas activated during and the ECoG features best associated with a syllable repetition task, which paves the way for more complex speech decoding.

A study by Zhang *et al*. (2012) used a sentence-level approach to classify ECoG high gamma responses obtained from the posterior portion of the inferior frontal gyrus during the production of two eight-character Chinese sentences with a 77.5% accuracy of classification (chance level 50%) [53]. Dynamic time warping was used to align the ECoG responses with the onset of sentence articulation and identify temporal activation patterns, or ECoG response templates, for the two sentences. Fisher linear discriminant analysis (LDA) was used for classification of the two sentences based on the time-warped ECoG responses, which was then evaluated using a leave-one-out cross validation technique. This study demonstrates the discriminability of high gamma activity recorded during speech at the sentence level, which is an important advance toward a neuroprosthesis for fluent, conversational speech. The studies discussed to this point have attempted to decode semantic categories, words, vowels, consonants and sentences in articulated speech from ECoG activity, and have shown preliminary success in speech decoding. However, the success rates for these decoding techniques are still low for a speech prosthesis capable of operating in natural environments where both decoding speed and accuracy are of importance.

One possible way to improve the information rate of speech decoding may be to predict the smallest identifiable components of speech, called phonemes, which can be sequenced together to form more complex productions. A study by Blakely *et al*. (2008) used micro-ECoG grids to successfully classify a set of four phonemes, in a pair-wise fashion, using ECoG high-gamma power [25]. The ten best channels for classification of each of the phoneme pairs were found using a correlation-based feature selection technique, followed by a binary classification with a linear support vector machine, which was then validated using a 4-fold nested cross-validation. The study found that different locations on the cortex were specific to classification of particular phoneme pairs, thus demonstrating the spatial separation of phoneme representation in the human brain. Using the best set of electrodes for each phoneme pair, accuracies as high as 75% for classification of the /ra/ vs /la/ pair and 70% for the /ba/ vs /wa/ pair (chance level 50% for both), were achieved. Extending these possibilities further, a recent study by Mugler *et al*. (2014) investigated a technique to decode the entire set of phonemes in American English using ECoG recordings from four subjects while they produced words from the modified rhyme test. This test consists of 300 words with similar frequencies of phoneme occurrence as found in the English language [54]. ECoG feature selection was performed on the time-frequency features (short time Fourier Transform features in the mu, beta and high gamma frequency bands) using an ANOVA and selecting features as those with the lowest p-values. These features were then used to classify phonemes using linear discriminant analysis technique followed by a ten-fold cross-validation to evaluate the classification performance. For the subject with the best performance, 36.1% of all consonant phonemes (chance level 7.4%) and 23.9% of all vowel phonemes (chance level 12.9%) were correctly classified, with a classification rate as high as 63% for a single phoneme. Another study by Leuthardt *et al*. (2011) was able to successfully classify the production of two phonemes based on the squared correlation of the ECoG high gamma power during overt speech production [27]. This was an online study where two subjects produced two different phonemes to control a one-dimensional cursor on the computer screen. The cursor was controlled using a weighted summed value, based on the decoded phoneme (e.g., one phoneme moved the cursor right, the other left). Classification rates of 76% and 91% were achieved for the two subjects (chance level 46.2%).

Decoding of imagined speech production

Because the primary goal of a speech neuroprosthesis is to restore communication to those who are able to produce little or no normal verbal communication, it is vital to demonstrate that imagined or attempted speech can be accurately decoded from brain activity. Pei *et al*. (2011) examined ECoG recordings to classify vowels and consonant pairings during covert word repetition, achieving classification accuracies as high as 43% for imagined vowel classification and 46% for consonant pairs (chance level 25% for both) in imagined speech [51]. This was done using the same technique for feature selection (MRMR), classification (Naive Bayes Classifier), and evaluation (10-fold cross-validation) as used for the decoding of overt vowels and consonant pairs. This was one of the first studies to demonstrate the possibility of classifying different vowels and consonants embedded in imagined words directly from brain signals. Furthermore, the decoding results were similar for both actual and imagined speech, which provides evidence that imagined speech can also be decoded from neural activity. Leuthardt *et al*. (2011) also investigated online control of a one-dimensional cursor using ECoG high-gamma power for an imagined phoneme versus rest task in two subjects (e.g., imagination of a phoneme moved the cursor right, rest moved it left) [27]. The feature selection and classification techniques used were consistent for the overt and covert phoneme pair identification (discussed in the 'Decoding of Overt Speech Production' section of this article). This resulted in closed-loop classification accuracies above 69% (chance level 42.6%) for both the overt and covert conditions, with the overt condition yielding a performance as high as 91%.

A recent study by Martin *et al*. (2014) explored the possibility of predicting spectro-temporal components of imagined speech from ECoG high gamma power, in a manner similar to overt speech [55]. In this study, subjects read aloud (overt condition) and imagined reading (covert condition) short stories that scrolled across the computer screen. Neural decoding models were then developed for the overt speech condition in order to predict two speech feature representations: (1) a spectrogram-based feature, which is a time-varying speech amplitude envelope at different acoustic frequencies, and (2) a modulation-based feature, which is a non-linear transformation of the spectrogram. Linear decoding models were developed in order to predict the two speech representations from ECoG high-gamma activity for the overt condition. These models were then applied to predict the speech representations during the covert condition. Dynamic time warping was used to align the reconstructed covert speech representations to the actual overt representations. The correlation coefficients between the actual and predicted speech representations for the overt condition, and between the time-warped actual and predicted speech representations for the covert condition, were used to evaluate the models. The reconstruction correlation was found to be statistically significant in all the subjects for the overt condition. For the covert condition, the predictions were statistically significant when compared to the baseline condition. This indicates that auditory representations of imagined speech can be reconstructed from models developed for actual speech, showing that both overt and covert conditions share a common neural basis.

Decoding of speech perception

Other studies have investigated the possibility of decoding perceived speech directly from cortical recordings. A study by Zavaglia *et al*. (2012) analyzed auditory features to build a forward model of the ECoG responses corresponding to word and acoustically matched non-word stimuli presented to the subject [56]. This is done by using a weakly-coupled oscillator model of transient synchronization (WCO-TS). The WCO-TS uses the auditory stimulus being presented to the subject as the input and utilizes the serial nature of word processing in the human cortex, as demonstrated in [46], in order to predict the ECoG gamma activity corresponding to incoming auditory stimuli. Although this is inverse to the process of speech decoding, i.e. utilizing speech features to predict neural information, it provides useful information that may be analyzed to build a direct model for the prediction of speech features from ECoG. This study also identified a set of speech features called the "occurrence time" features which were found to outperform standard cepstral features typically used in speech recognition, especially in noisy recognition environments. These occurrence time features correspond to the occurrence of peaks in specific speech frequency bands and may be useful in future decoding efforts.

Chang *et al*. (2010) measured ECoG activity in the posterior STG using a high density micro-ECoG grid during the presentation of three consonant-vowel syllables [26]. The study found that an acoustically varying speech stimulus is transformed into distinct phoneme categories in the human cortex. Using this information, they were able to classify three consonant-vowel syllables from the ECoG signals recorded across the posterior STG. The dissimilarities between the neuronal response patterns were determined using a multivariate pattern classifier which uses L-1 norm regularized logistic regression, whose classification measures were used to construct a confusion matrix for each time interval. Multi-dimensional scaling of this confusion matrix and k-means were then used to classify the neuronal responses into three categories that corresponded to the three phonemes. The results from this study indicate that the posterior STG performs a critical role in the phonological processing and categorization of perceived speech.

A significant contribution for perceived speech prediction was made by Pasley *et al*. (2012), in a study that examined ECoG recordings from the STG to reconstruct the speech spectrogram of aurally presented words and sentences [57]. Two representations for the perceived speech were found, similar to those used in [55], i.e. a spectrogram-based representation and a non-linear modulation-based representation. Linear neural decoding models were then developed which used the ECoG high-gamma power to predict these two speech representations, leading to linear and non-linear models respectively. It was found that slow and moderate time modulations in the speech, such as syllable rate, were reconstructed well using the linear model, i.e., these modulations are well-represented with the spectrogram-based representation. Fast temporal modulations, such as syllable onsets and offsets, could be better predicted using the non-linear model, i.e. these modulations are well-represented with the modulation-based representation. The fidelity of the reconstructions of the spectrogram were sufficiently accurate to identify individual words directly from the reconstructed spectrogram-based speech representations using a simple spectrogram matching algorithm, leading to a median word identification percentile rank of 0.89 for 47 words (chance level 0.50).

FUTURE DIRECTIONS TOWARDS THE DEVELOPMENT OF A PRACTICAL SPEECH NEUROPROSTHETIC

These collections of studies have identified important neural correlates associated with speech production and perception. Decoding models have also been successfully developed that are capable of predicting the essential components of vocal communication, namely the production and perception of speech and language directly from cortical activity. ECoG studies have been especially informative in their ability to provide a spatiotemporal characterization of the neural correlates of speech planning, production and perception in the human cortex. These studies have specifically focused on the ECoG gamma-band power recorded over language-related cortical areas, which is significantly correlated with speech processing and is useful for speech decoding algorithms. However, the performance of ECoG-based speech decoders are not nearly as robust as needed for practical ECoG-based speech reconstruction, which is the ultimate aim of this type of research. One focus of future ECoG-based speech studies should include improving computational models of neurological speech processing for more accurate decoding. New models will benefit from continued research on the development of more advanced signal processing and ECoG electrode design to capture the recorded signals with higher fidelity. Most of the studies described in this review have implemented relatively simplistic linear approaches to characterize and predict speech components from cortical activity. However, in reality, it is likely that the relationship between ECoG activity and the various speech representations of interest are highly non-linear and dynamic. Therefore, more sophisticated models based on improved signal acquisition capable of capturing such non-linear relationships need to be developed to achieve a more transparent and practical ECoG based speech decoder.

The majority of speech and language ECoG studies have been performed using relatively discontinuous speech production and listening tasks, such as cued word repetition tasks. These studies are critical for identifying baseline neurological activation during speech, but are not fully representative of fluent, conversational speech. A characterization of the neural correlates associated with continuous and spontaneous speech production and perception may provide the supplementary information needed to develop more advanced models. The ability to decode perceived speech and articulatory commands in continuous and fluent communication will represent a fundamental improvement in the potential impact of a neural prosthesis for speech. Natural verbal communication often takes place in the presence of background speech or environmental noise. The perception and comprehension of speech in background noise requires additional verbal working memory and attentional resources to process the target stream and segregate it from competing background noise [59-61]. In addition, the auditory-motor feedback loop active during vocalization has been shown to affect speech production [47, 49]. However, the potential interference of background noise on the feedback loop during communication is not well understood, particularly at low signal-to-noise ratios, and certainly not included in current decoding algorithms. Future ECoG speech studies will need to determine the contribution and coordination of both specific speech and non-speech regions to the production and processing of speech in the presence of different levels of background noise.

Furthermore, most of the ECoG-based speech decoding studies have focused on decoding speech features such as the envelope, words, phonemes, vowels and consonants from ECoG activity. However, modern real-time speech processors and automatic speech recognition (ASR) systems employ many other speech features, such as the formant frequencies, linear predictive coding coefficients, and mel frequency cepstral coefficients, among others [62]. While some studies have used formants for ECoG based speech decoding [26], most other speech representations used in ASR have not been investigated with regard to their relationship to cortical activity. The decoding of these speech representations directly from ECoG activity is a practical next step, and the results could then be used in speech synthesis and recognition systems. Along these lines, the incorporation of language models and other ASR techniques should be examined to further improve performance. This research would develop a natural extension of ASR to neurological data, and provide a step toward neural speech prostheses.

Several of the ECoG studies discussed here use micro-ECoG grids, i.e. ECoG grids with more densely spaced electrodes than regular ECoG grids, in order to decode speech [24-27]. These micro-ECoG grids offer higher spatial resolutions than regular ECoG grids and hence, can be beneficial for localization and decoding of speech components more effectively than standard clinical ECoG grid spacing. Other recent speech studies have analyzed intracortical electrode recordings that penetrate the cortex in order to decode various components of speech, such as imagined vowels and phonemes, directly from cortical activity [63, 64]. However, human research using micro-ECoG and intracortical electrodes has been limited due to the lack of clinical applicability of these electrodes. Nevertheless, it can be argued that the addition of micro-ECoG to standard clinical grids poses insignificant risk to the patient, and that it may even benefit seizure localization and cortical mapping. For these reasons, it is expected that micro-ECoG will continue to gain clinical acceptance, which will in turn benefit future language decoding and general neuroscience studies.

The ability to modulate pitch, intensity, speaking rate, syllabic stress and rhythm is an important part of communication. These variables can combine to convey emotional affect and linguistic information beyond standard speech production. Ideally, a practical neuroprosthesis for communication would decode the combination of speech output, intended emotion, motor plans for facial expressions, and intended patterns of intonation and stress together from acquired neural signals for natural speech output. In addition, a practical prosthesis should be able to distinguish between intended speech output and "inner speech" (e.g., differences between overt and covert speech) to allow potential users a way to monitor and filter their speech output. This description of a neural prosthesis for communication is far beyond the capabilities of any current speech decoding efforts. In addition, some intended users of speech and language neuroprosthetics may present with varying cognitive and motor abilities making it difficult for them to control such a complex neuroprosthetic device. However, a natural communication prosthesis with a direct link to the brain should possess these abilities to offer users the most complete communication experience possible.

Finally, while studies that analyze ECoG signal recordings have provided invaluable information toward understanding the complex processes involved in speech production and auditory processing, these studies have examined recordings obtained over a short duration in a medical setting. Typically, recordings are obtained over the course of a few minutes to several days while the subjects are stationary, confined to a hospital bed, and interacting with one speaker. In contrast, natural speech communication occurs in the presence of many additional factors, including visual scene changes and additional motor movements during activities such as walking, nonverbal communication including gestures and eye gaze maintenance, interaction with multiple and varying communication partners, and conversation maintenance and repair after interruption or misunderstanding during lengthy communication exchanges. In order to develop a neural prosthesis for speech and language that successfully functions in natural settings, these factors must be examined in future studies. Chronically implanted electrodes would allow for ECoG recordings in natural communication environments and represent one way to investigate these factors in communication. Few studies have been conducted on the safety and efficacy of long-term electrode implantation in humans, or the long-term placement of ECoG grids. However, early evidence from recent studies of the stability of long-term impedance in chronic subdural electrodes concluded that impedance was stable up to one year after implantation [29, 30]. In addition, these studies reported few adverse effects resulting from chronically implanted electrodes. This highlights the potential feasibility of the use of chronic subdural electrodes for future speech-based neuroprosthetics.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation (1064912, 1451028) and the National Institute on Deafness and Other Communication Disorders (R03-DC011304).

CONFLICT OF INTEREST STATEMENTS

Chakrabarti S declares that she has no conflict of interest in relation to the work in this article. Sandberg HM declares that she has no conflict of interest in relation to the work in this article. Brumberg JS declares that he has no conflict of interest in relation to the work in this article. Krusienski DJ declares that he has no conflict of interest in relation to the work in this article.

REFERENCES

[1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Clin Neurophysiol. 2002; 113(6):767-91.
[2] Ficke RC. Digest of data on persons with disabilities. Washington, DC: National Institute on Disability and Rehabilitation Research. 1992.
[3] Pasley BN, Knight RT. Decoding speech for understanding and treating aphasia. Prog Brain Res. 2013; 207:435-56.
[4] 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):816-47.
[5] Pei X, Hill J, Schalk G. Silent communication: toward using brain signals. IEEE Pulse. 2012; 3(1):43-6.
[6] Ojemann GA. Cortical organization of language. J Neurosci. 1991; 11(8):2281-7.
[7] Broca P. Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche du cerveau. Bull Soc Anthropol. 1861; 2:235-8.
[8] Wernicke C. Der aphasische symptomenkomplex. Springer Berlin Heidelberg: 1974.
[9] Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci. 2007; 8(5):393-402.
[10] Hickok G. Computational neuroanatomy of speech production. Nat Rev Neurosci. 2012; 13(2):135-45.
[11] Guenther FH, Ghosh SS, Tourville JA. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 2006; 96(3):280-301.
[12] Price CJ, Wise RJ, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RS, Friston KJ. Hearing and saying the functional neuro-anatomy of auditory word processing. Brain. 1996; 119(3):919-31.
[13] Price CJ. The anatomy of language: contributions from functional neuroimaging. J Anat. 2000; 197(3):335-59.
[14] Fiez JA, Petersen SE. Neuroimaging studies of word reading. Proc Natl Acad Sci USA. 1998; 95(3):914-21.
[15] Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 2000; 10(5):512-28.
[16] Talavage TM, Gonzalez-Castillo J, Scott SK. Auditory neuroimaging with fMRI and PET. Hear Res. 2014; 307:4-15.
[17] Ganushchak LY, Christoffels IK, Schiller NO. The use of electroencephalography in language production research: a review. Front Psychol. 2011; 2(208):1-6.
[18] Sanders LD, Neville HJ. An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. Brain Res Cogn Brain Res. 2003; 15(3):228-40.
[19] Hagoort P, Brown CM. ERP effects of listening to speech: semantic ERP effects. Neuropsychologia. 2000; 38(11):1518-30.
[20] Indefrey P, Levelt WJ. The spatial and temporal signatures of word production components. Cognition. 2004; 92(1-2):101-44.
[21] Leuthardt EC, Pei XM, Breshears J, Gaona C, Sharma M, Freudenberg Z, Barbour D, Schalk G. Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. Front Hum Neurosci. 2012; 6:99.
[22] Palmini A. The concept of the epileptogenic zone: a modern look at Penfield and Jasper's views on the role of interictal spikes. Epileptic Disord. 2006; 8 Suppl 2:S10-5.
[23] Schalk G, Leuthardt EC. Brain-computer interfaces using electrocorticographic signals. IEEE Rev Biomed Eng. 2011; 4:140-54.
[24] 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(5):056007.
[25] Blakely T, Miller KJ, Rao RP, Holmes MD, Ojemann JG. Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. Conf Proc IEEE Eng Med Biol Soc. 2008; 2008:4964-7.
[26] Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM,

Knight RT. Categorical speech representation in human superior temporal gyrus. Nat Neurosci. 2010; 13(11):1428-32.
[27] Leuthardt EC, Gaona C, Sharma M, Szrama N, Roland J, Freudenberg Z, Solis J, Breshears J, Schalk G. Using the electrocorticographic speech network to control a brain–computer interface in humans. J Neural Eng. 2011; 8(3):036004.
[28] Schwartz AB, Cui XT, Weber DJ, Moran DW. Brain-controlled interfaces: movement restoration with neural prosthetics. Neuron. 2006; 52(1):205-20.
[29] Sillay KA, Rutecki P, Cicora K, Worrell G, Drazkowski J, Shih JJ, Sharan AD, Morrell MJ, Williams J, Wingeier B. Long-term measurement of impedance in chronically implanted depth and subdural electrodes during responsive neurostimulation in humans. Brain Stimul. 2013; 6(5):718-26.
[30] Wu C, Evans JJ, Skidmore C, Sperling MR, Sharan AD. Impedance variations over time for a closed-loop neurostimulation device: early experience with chronically implanted electrodes. Neuromodulation. 2013; 16(1):46-50.
[31] Crone NE, Sinai A, Korzeniewska A. High-frequency gamma oscillations and human brain mapping with electrocorticography. Prog Brain Res. 2006; 159:275-95.
[32] Leuthardt EC, Schalk G, Wolpaw JR, Ojemann JG, Moran DW. A brain-computer interface using electrocorticographic signals in humans. J Neural Eng. 2004; 1(2):63-71.
[33] Yanagisawa T, Hirata M, Saitoh Y, Goto T, Kishima H, Fukuma R, Yokoi H, Kamitani Y, Yoshimine T. Real-time control of a prosthetic hand using human electrocorticography signals. J Neurosurg. 2011; 114(6):1715-22.
[34] Schalk G, Miller KJ, Anderson NR, Wilson JA, Smyth MD, Ojemann JG, Moran DW, Wolpaw JR, Leuthardt EC. Two-dimensional movement control using electrocorticographic signals in humans. J Neural Eng. 2008; 5(1):75-84.
[35] Hinterberger T, Widman G, Lal TN, Hill J, Tangermann M, Rosenstiel W, Schölkopf B, Elger C, Birbaumer N. Voluntary brain regulation and communication with electrocorticogram signals. Epilepsy Behav. 2008; 13(2):300-6.
[36] Crone NE, Boatman D, Gordon B, Hao L. Induced electrocorticographic gamma activity during auditory perception. Clin Neurophysiol. 2001; 112(4):565-82.
[37] 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. 2011; 54(4):2960-72.
[38] Crone NE, Hao L, Hart J Jr, Boatman D, Lesser RP, Irizarry R, Gordon B. Electrocorticographic gamma activity during word production in spoken and sign language. Neurology. 2001; 57(11):2045-53.
[39] Sinai A, Bowers CW, Crainiceanu CM, Boatman D, Gordon B, Lesser RP, Lenz FA, Crone NE. Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming. Brain. 2005; 128(7):1556-70.
[40] Edwards E, Soltani M, Deouell LY, Berger MS, Knight RT. High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. J Neurophysiol. 2005; 94(6):4269-80.
[41] Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. Nature. 2013; 495(7441):327-32.
[42] Miller KJ, Abel TJ, Hebb AO, Ojemann JG. Rapid online language mapping with electrocorticography. J Neurosurg Pediatr. 2011; 7(5):482-90.
[43] Kubanek J, Brunner P, Gunduz A, Poeppel D, Schalk G. The tracking of speech envelope in the human cortex. PLoS One. 2013; 8(1):e53398.
[44] 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(1):377-86.
[45] 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(49):15564-74.
[46] 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(1):185-96.
[47] Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF. Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc Natl Acad Sci USA. 2013; 110(7):2653-8.
[48] Towle VL, Yoon HA, Castelle M, Edgar JC, Biassou NM, Frim DM, Spire JP, Kohrman MH. ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain. 2008; 131(8):2013-27.
[49] Greenlee JD, Jackson AW, Chen F, Larson CR, Oya H, Kawasaki H, Chen H, Howard MA 3rd. Human auditory cortical activation during self-vocalization. PLoS One. 2011; 6(3):e14744.
[50] Wang W, Degenhart AD, Sudre GP, Pomerleau DA, Tyler-Kabara EC. Decoding semantic information from human electrocorticographic (ECoG) signals. Conf Proc IEEE Eng Med Biol Soc. 2011; 2011:6294-8.
[51] Pei X, Barbour DL, Leuthardt EC, Schalk G. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. J Neural Eng. 2011; 8(4):046028.
[52] Kanas VG, Mporas I, Benz HL, Sgarbas KN, Bezerianos A, Crone NE. Joint spatial-spectral feature space clustering for speech activity detection from ECoG signals. IEEE Trans Biomed Eng. 2014; 61(4):1241-50.
[53] Zhang D, Gong E, Wu W, Lin J, Zhou W, Hong B. Spoken sentences decoding based on intracranial high gamma response using dynamic time warping. Conf Proc IEEE Eng Med Biol Soc. 2012; 2012:3292-5.
[54] Mugler EM, Patton JL, Flint RD, Wright ZA, Schuele SU, Rosenow J, Shih JJ, Krusienski DJ, Slutzky MW. Direct classification of all American English phonemes using signals from functional speech motor cortex. J Neural Eng. 2014; 11(3):035015.
[55] Martin S, Brunner P, Holdgraf C, Heinze HJ, Crone NE, Rieger J, Schalk G, Knight RT, Pasley BN. Decoding spectrotemporal features of overt and covert speech from the human cortex. Front Neuroeng. 2014; 7:14.
[56] Zavaglia M, Canolty RT, Schofield TM, Leff AP, Ursino M, Knight RT, Penny WD. A dynamical pattern recognition model of gamma activity in auditory cortex. Neural Netw. 2012; 28:1-14.
[57] 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.
[58] Behroozmand R, Larson CR. Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci. 2011; 12:54.
[59] Parbery-Clark A, Strait DL, Anderson S, Hittner E, Kraus N. Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. PLoS One. 2011; 6(5):e18082.
[60] Heinrich A, Schneider BA, Craik FI. Investigating the influence of continuous babble on auditory short-term memory performance. Q J Exp Psychol. 2008; 61(5):735-51.
[61] Pichora-Fuller MK. Audition and cognition: What audiologists

need to know about listening. In: Palmer C, Seewald R, editors. Hearing Care for Adults. Stäfa, Switzerland: Phonak; 2007. pp 71-85.
[62] Deng L, O'Shaughnessy D. Speech processing: a dynamic and optimization-oriented approach. CRC Press; 2003.
[63] 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(12):e8218.
[64] Brumberg JS, Wright EJ, Andreasen DS, Guenther FH, Kennedy PR. Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. Front Neurosci. 2011; 5:65.