



The Potential for a Speech Brain–Computer Interface Using Chronic Electrocochicography

Qinwan Rabbani¹ · Griffin Milsap² · Nathan E. Crone³

Published online: 7 January 2019
© The American Society for Experimental NeuroTherapeutics, Inc. 2019

Abstract

A brain–computer interface (BCI) is a technology that uses neural features to restore or augment the capabilities of its user. A BCI for speech would enable communication in real time via neural correlates of attempted or imagined speech. Such a technology would potentially restore communication and improve quality of life for locked-in patients and other patients with severe communication disorders. There have been many recent developments in neural decoders, neural feature extraction, and brain recording modalities facilitating BCI for the control of prosthetics and in automatic speech recognition (ASR). Indeed, ASR and related fields have developed significantly over the past years, and many lend many insights into the requirements, goals, and strategies for speech BCI. Neural speech decoding is a comparatively new field but has shown much promise with recent studies demonstrating semantic, auditory, and articulatory decoding using electrocochicography (ECoG) and other neural recording modalities. Because the neural representations for speech and language are widely distributed over cortical regions spanning the frontal, parietal, and temporal lobes, the mesoscopic scale of population activity captured by ECoG surface electrode arrays may have distinct advantages for speech BCI, in contrast to the advantages of microelectrode arrays for upper-limb BCI. Nevertheless, there remain many challenges for the translation of speech BCIs to clinical populations. This review discusses and outlines the current state-of-the-art for speech BCI and explores what a speech BCI using chronic ECoG might entail.

Key Words Electrocochicography (ECoG) · Brain–computer interface (BCI) · Neural speech decoding · Automatic speech recognition (ASR) · Locked-in syndrome (LIS) · Communication

Background

Brain–Computer Interfaces and Speech BCI

A brain–computer interface (BCI) is a computational device or platform that takes an input signal from the brain and decodes some type of information from it, in order to control a physical device or software interface and/or facilitate sensori-

motor or cognitive restoration or augmentation. Most broadly, a BCI serves as an interface between the human brain and technology. The literature refers to BCI using a number of synonyms, including brain–machine interface (BMI) and neural prosthetic.

A speech BCI is a device that produces some form of speech output (e.g., words, sentences, synthesized speech) from a measure of the user’s brain signals. Such a technology could also serve as a useful control signal for human–computer interactions, such as navigating the internet, controlling home automation, or communicating with a personal assistant. Speech BCI is in early developmental stages with two primary tracks: interfaces using noninvasive recording modalities and interfaces using implanted electrodes. In the past, state-of-the-art BCIs from the latter track of development were not convincing enough to outweigh the risks of implantation, even for patient populations that could benefit most. However, we believe that the field is now at a critical turning point, in which scientific advances justify the risks of an implanted speech BCI.

✉ Qinwan Rabbani
qrabbani@jhu.edu

¹ Department of Electrical Engineering, The Johns Hopkins University Whiting School of Engineering, Baltimore, MD, USA

² Department of Biomedical Engineering, The Johns Hopkins University School of Medicine, Baltimore, MD, USA

³ Department of Neurology, The Johns Hopkins University School of Medicine, Baltimore, MD, USA

Target Population

In its current state, implanting a BCI would be reasonable only in a target population for whom the potential benefits would outweigh the risks of brain surgery. A variety of patient populations have expressive speech impairments and could potentially benefit from such a device. The largest populations arguably consist of patients with primary progressive aphasia due to stroke or neurodegenerative speech disorders, but these disorders pose a far greater challenge due to lesions of cortical areas that would be important for a speech BCI [1]. On the other hand, patients with locked-in syndrome (LIS) are aware and conscious and have intact language cortices, but are unable to communicate due to severe paralysis of the limbs, face, and muscles of articulation, despite remaining otherwise cognitively intact [2]. LIS can arise due to a variety of neurological causes, such as traumatic, neoplastic, inflammatory, or metabolic lesions of the brainstem or neuromuscular disorders like amyotrophic lateral sclerosis (commonly known as ALS) [2]. For many individuals with LIS, their only means of communication is through limited eye movements and blinking; however, for those with total locked-in syndrome, even this type of communication may be unreliable [2]. Those who are not yet in a totally locked-in state can often communicate through slow and labored speech, through the use of gestures or eye blinks or by focusing their gaze on specific letters and symbols on communication boards to signal their intention to a caretaker [2]. More high-tech solutions like voice amplification devices and eye-tracking devices may also assist in communication [2]. An ideal speech–BCI system with sufficiently high performance would provide an alternative to these systems, allowing those with LIS to communicate more effectively and efficiently without the need for a caretaker to initiate a conversation or to speak for them. Such a technology could empower these individuals and dramatically improve their quality of life.

Neural Bases of Speech Production

Despite significant progress over the past decade in machine learning and automatic speech recognition (ASR), there is still much to learn about the neurobiological mechanisms of speech production. Historic studies using functional lesioning and a variety of neuroimaging techniques have mapped the brain areas responsible for processing different aspects of speech and the general dynamics of how they interact [3]. Indefrey [3] has described one of the most comprehensive models of these dynamics to date based on a variety of imaging studies. Indefrey's [3] model separates speech production into 6 distinct stages: conceptual preparation, lemma retrieval, phonological code retrieval, phonological encoding, phonetic encoding, and articulation (see Fig. 1). Most simply, this stream of processes can be thought of as the serial transformation of speech information along a continuous “speech pathway.” In natural

speech, however, the transformation of information along this pathway is in fact cascaded and not strictly sequential, with substantial temporal overlap, parallel processing, and feedback in the cortical networks responsible for the aforementioned transformations [3]. Nevertheless, the functional-anatomic compartmentalization of the neural speech pathway suggests that neural speech decoding could intercept this stream at many different points along the pathway with different tradeoffs.

Electrooculography in BCI

Not all brain recording modalities are ideal for real-time BCI. Noninvasive magnetic imaging modalities like functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS) can interrogate the majority of the brain surface at good spatial resolution, albeit through an indirect, delayed measure of neural activity, the hemodynamic or blood-oxygen-level dependent (BOLD) response (with hemodynamic lag), with insufficient temporal resolution to match real-time speech synthesis [4, 5]. For this reason, these modalities are ill-suited for real-time BCI control. Although magnetoencephalography (MEG) has good spatial resolution and excellent temporal resolution, its neural source localization can be challenging [6], and its need for electromagnetic shielding [7] poses significant challenges in deploying MEG for real-world BCI.

The current gold standards for BCI, electroencephalography (EEG) and microelectrode arrays (MEAs), have shown much success in the literature, but they too have drawbacks. EEG can interrogate most of the brain surface with excellent temporal resolution but has poor spatial resolution and poor signal quality (particularly for high-frequency activity) largely due to the distance of the electrodes from the electrical source(s) in cortex [8–10]. Nonetheless, EEG has been widely adopted due to its portability and its ability to resolve changes in overall brain states for simple-but-effective BCI designs [11]. On the other hand, MEAs record activity from single neurons with excellent temporal resolution but sample this activity from a very small patch of cortex [12]. MEA needle electrodes physically penetrate the brain and elicit tissue responses that affect signal stability over the long-term due to scarring and neuronal cell death [13, 14]. Nonetheless, MEAs facilitate effective BCIs for motor prosthetic control [15–17]. However, we posit that their limited spatial coverage is insufficient for neural speech decoding, in which sampling neural activity across a large speech network is likely necessary for success.

ECoG electrodes are typically composed of platinum–iridium discs that are embedded in silastic sheets that are surgically implanted directly beneath the dura (subdurally), though they may also be implanted above it (epidurally) [18]. These electrodes present a potentially ideal compromise between the drawbacks of the other brain recording modalities. Figure 2 shows a typical ECoG array. Like EEG, ECoG

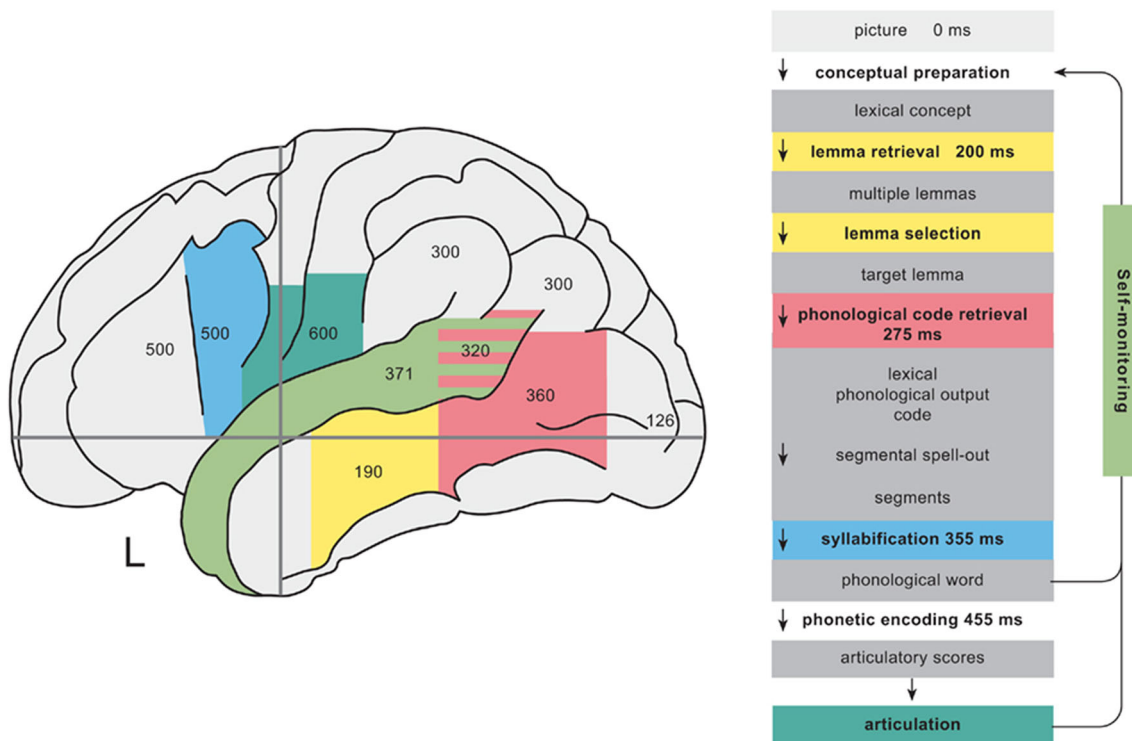


Fig. 1 Stages of speech production according to Indefrey’s [3] model. The color-coded table to the right summarizes the stages of processing in speech production and time course in response to an image stimulus. The

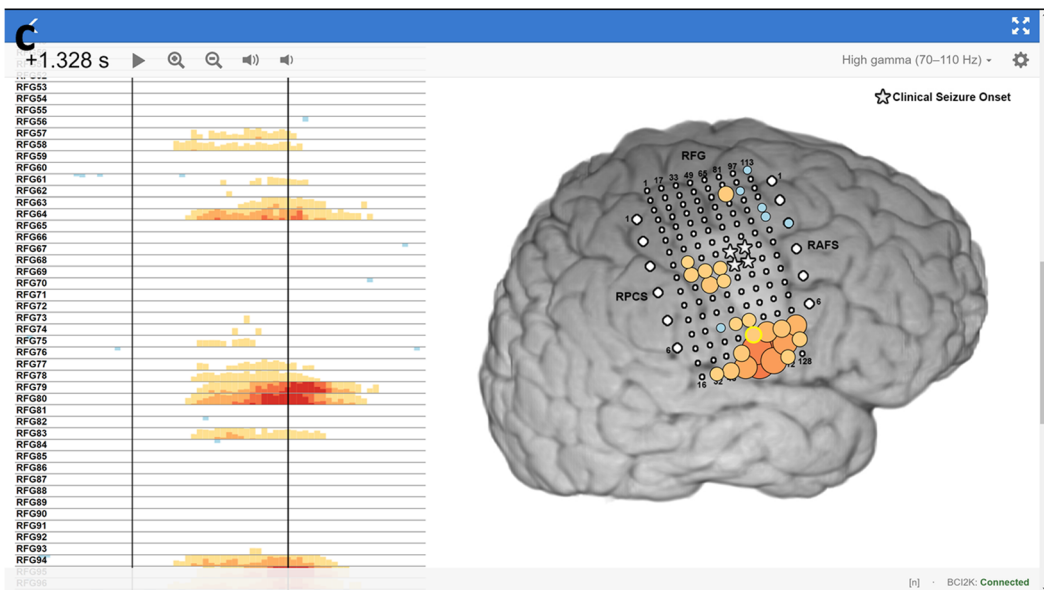
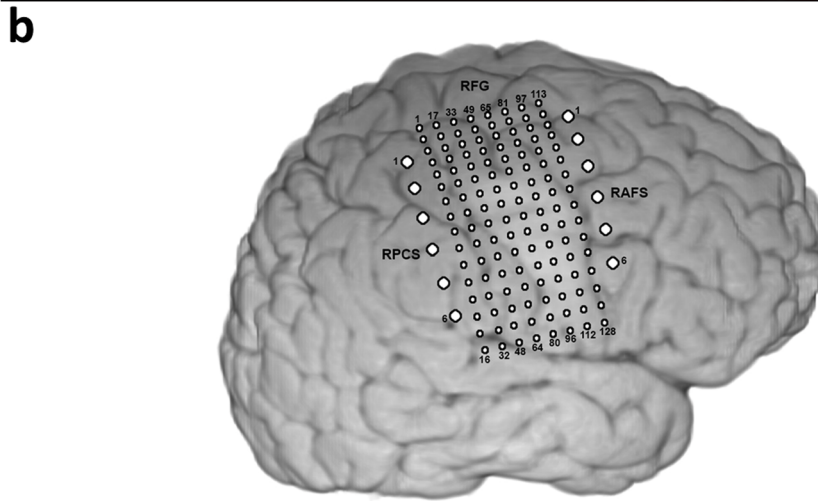
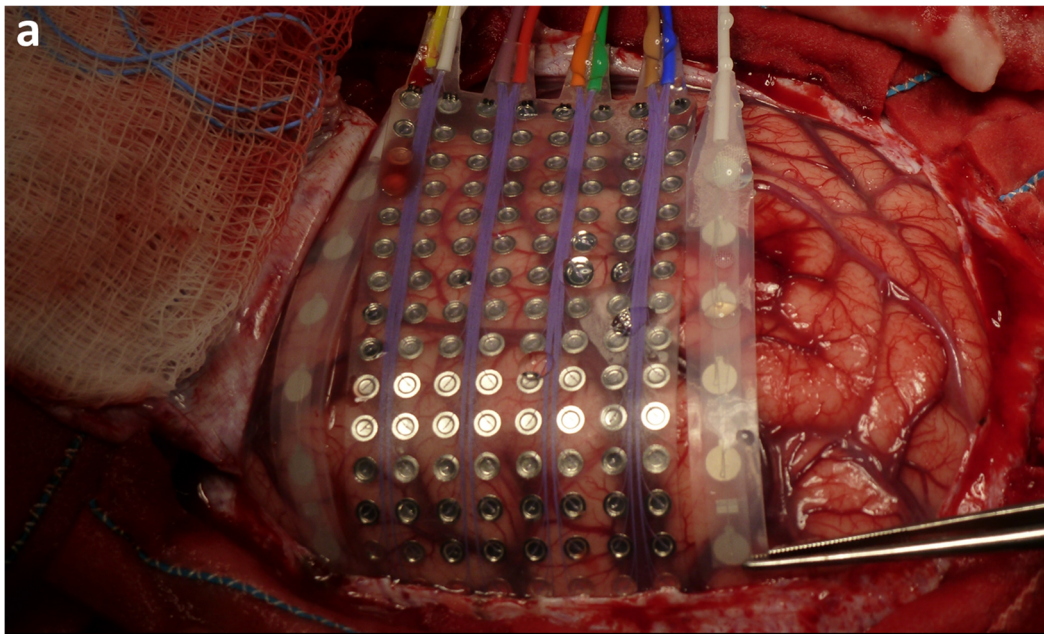
color-coded image to the left displays regions implicated in each stage of processing along with peak activation times observed experimentally in milliseconds [3]. Figure reused with permission from Ref. [3]

measures the summed local synaptic field potentials (LFPs) of underlying neuronal populations [18], a direct measure of neural activity inaccessible to fMRI [4]. Because the electrodes are in direct contact with the cortical surface, source localization is trivial, and signal quality is greatly enhanced, especially for high-frequency activity [18]. ECoG electrodes can also feasibly be combined into grids and strips consisting of dozens or hundreds of electrodes to sparsely cover large and/or distributed regions of the brain [18]. Though the ECoG arrays used in epilepsy seizure monitoring typically have an exposed surface diameter of 2.3 mm and an interelectrode spacing of 10 mm [18], these arrays can also be manufactured with much smaller electrode dimensions and spacing going as low as 1 mm in diameter and in spacing [20]. Together, these allow ECoG to measure neural-population activity with unrivaled spatial and temporal resolution.

ECoG also offers advantages for decoding. It is likely that many degrees of freedom will be necessary to adequately model enough speech articulators simultaneously (e.g., one per articulator) or to adequately reconstruct the complexity of the speech waveform (e.g., one per spectrotemporal bin, etc.). Speech is distributed in many different parts of the brain, each playing an integral role in representing the different aspects of speech. Because ECoG is able to record activity from these multiple, highly distributed speech areas simultaneously with high quality, it may be more naturally attuned to providing the degrees of freedom necessary for high-performance decoding.

However, ECoG also has drawbacks. Its biggest drawback is its relative invasiveness, requiring burr holes or craniotomies to place the electrodes [18]. For this reason, ECoG has most often been used in drug-resistant epilepsy patients, as well as in tumor patients, undergoing resection [18]. Electrode array placement in these patients is dictated solely by medical considerations [18]. Nonetheless, one or more segments of the speech network may be covered by electrodes in some patients who volunteer for research exploring the feasibility of ECoG-based speech BCIs. Because this line of research is restricted to a relatively small clinical population with intact speech, its generalizability to the target population of patients with communication impairments is limited. However, this issue can be partially accounted for by removing electrodes contaminated by proximity to epileptic foci or tumors and through careful patient selection. Results in otherwise cognitively normal patients are more likely to generalize to a healthy population.

In spite of the aforementioned drawbacks, ECoG studies in patients undergoing resection continue to contribute significant scientific information about the feasibility of ECoG-based speech BCIs. Based on the cumulative experience of multiple such studies around the world, there is growing confidence in the potential benefits of ECoG-based speech BCIs to overcome their inherent risks in patients with severe communication disabilities.



◀ **Fig. 2** Subdural electrocorticography (ECoG). (a) Intraoperative photograph of surgical implantation of a high-density 128-channel grid (8×16 electrodes, 2-mm diameter, 5-mm spacing) flanked by two standard linear 1×8 electrode strips (4-mm diameter, 2.3-mm diameter exposed surface area, 10-mm spacing). (b) Surface rendering of brain reconstruction from registration between preoperative MRI and postoperative CT image. (c) Trial-averaged high gamma power changes (increases shown with red color scale, decreases shown with blue color scale) relative to prestimulus baseline during syllable reading are shown for each ECoG recording site (average reference) in a channel raster in WebFM [19]. The bright patch of cortex corresponds to a low-grade glioma in the precentral gyrus. The size and hue of the circles over the colored electrodes (c) correspond to the magnitude of relative high gamma power changes (increases shown with red color scale, decreases shown with blue color scale). Time is shown relative to cue onset (c) at 0 s (left vertical line) with high gamma power reflected on the brain at a single time point (right vertical line, 1.328 s). High gamma power increases (c) occur in a large area of the ventral sensorimotor cortex, as well as in a smaller, more dorsal area, in alignment with speech responses

Neural Speech Decoding

Types of Neural Speech Decoding

Speech can be examined through three major representations, semantic, auditory, and articulatory, each present at different points along the speech production pathway. Regarding semantics, speech can be represented by the underlying concepts or discrete words conveyed by that speech. Though the natural representation of semantics in the brain is not fully understood, various feature-encoding models exist that map semantic attributes of speech in high-dimensional space (see Fig. 3) [21]. When decoding speech production, semantics would likely be intercepted at one of the earliest stages of speech production, i.e., conceptual retrieval, in widely distributed regions in the brain, especially in the temporal lobe, as well as in the frontal, parietal, and occipital lobes [22]. Regarding audition, speech can be represented as its acoustic waveform and its corresponding spectrotemporal features (see Fig. 4). These features can be further segmented into phonemes, the fundamental sound units of words (e.g., the “ah,” “p,” and “l” sounds in “apple”), of which there are ~48 in American English [23]. Neural representations of phonemes or related acoustic representations could potentially be intercepted for decoding in one of the middle stages of speech production, i.e., phonological code retrieval or syllabification (or later if intercepted as an efference copy). This would likely require recordings from auditory association cortices within the temporal lobe, particularly along the lateral superior temporal gyrus (STG) [24]. Regarding articulation, speech can be represented as a sequence of phonation and articulatory gestures generated by the muscles and organs within the vocal tract and mouth including the tongue (see Fig. 5). Articulation has been well characterized by a source–filter model of speech production, in which the vocal cords (the source) vibrate to generate sound, which is then modulated by the positions of the

articulators (the filter) over time to generate phonetic sounds [25]. When decoding speech, this representation would likely be intercepted in one of the later stages of speech production, i.e., phonetic encoding or articulation in the speech-related motor integration areas of the brain, such as Broca’s area and the ventral sensorimotor cortex (vSMC) [27, 28].

The aforementioned representations could also be examined during overt (audible) or covert (silent or imagined) speech. Most work to date has focused on decoding overt speech, because it provides a ground-truth with which to align neural events with behavior, e.g., speech onset, offset and duration, and phoneme boundaries. Covert speech is difficult to decode because alignment of neural events with covert speech is ill-posed and because it is difficult to ensure subjects are actively participating. Semi-overt (or mimed) speech, in which articulation occurs without phonation, is rarely studied but can help isolate motor from auditory components of neural responses during speech production.

Each representation also offers unique benefits for speech BCI. The earlier the stage at which the speech representation is intercepted, the lower the latency at which the speech could be synthesized. Decoding the semantics of speech could be advantageous, as it could form the basis of an extremely versatile communication platform, potentially untethered to some of the normal constraints of language. However, because a compact natural representation of semantics is not known, and because decoding a potentially infinite number of semantic concepts is ill-constrained, progress has been limited in this area. On the other hand, decoding the auditory or articulatory aspects of speech could be advantageous due to their highly quantifiable spectrotemporal (acoustic) and gestural (articulatory) targets. In auditory decoding, spectrotemporal modeling of speech and auditory processing used in ASR, cochlear implant, and hearing research could potentially be borrowed to improve decoding using this approach. Similarly, in articulatory decoding, existing knowledge of somatotopic mapping of the speech articulators, work in vocal tract modeling [29], and articulatory-to-acoustic inversion [30] for speech synthesis and highly successful approaches from motor BCIs with MEAs could potentially be borrowed to improve decoding using this approach. Further, although work in covert neural speech decoding is limited, this may be less problematic for patients with severe neuromuscular weakness, in whom attempts to move may generate more neural activity in primary motor areas than would be generated by able-bodied subjects imagining the same movements [31, 32].

Semantic Decoding

Semantic decoding is a relatively new area in ECoG research, but key findings suggest its potential utility for speech BCI. The first relevant study by Liu et al. [33] was not intended to study semantics, but to study visual object recognition. In this

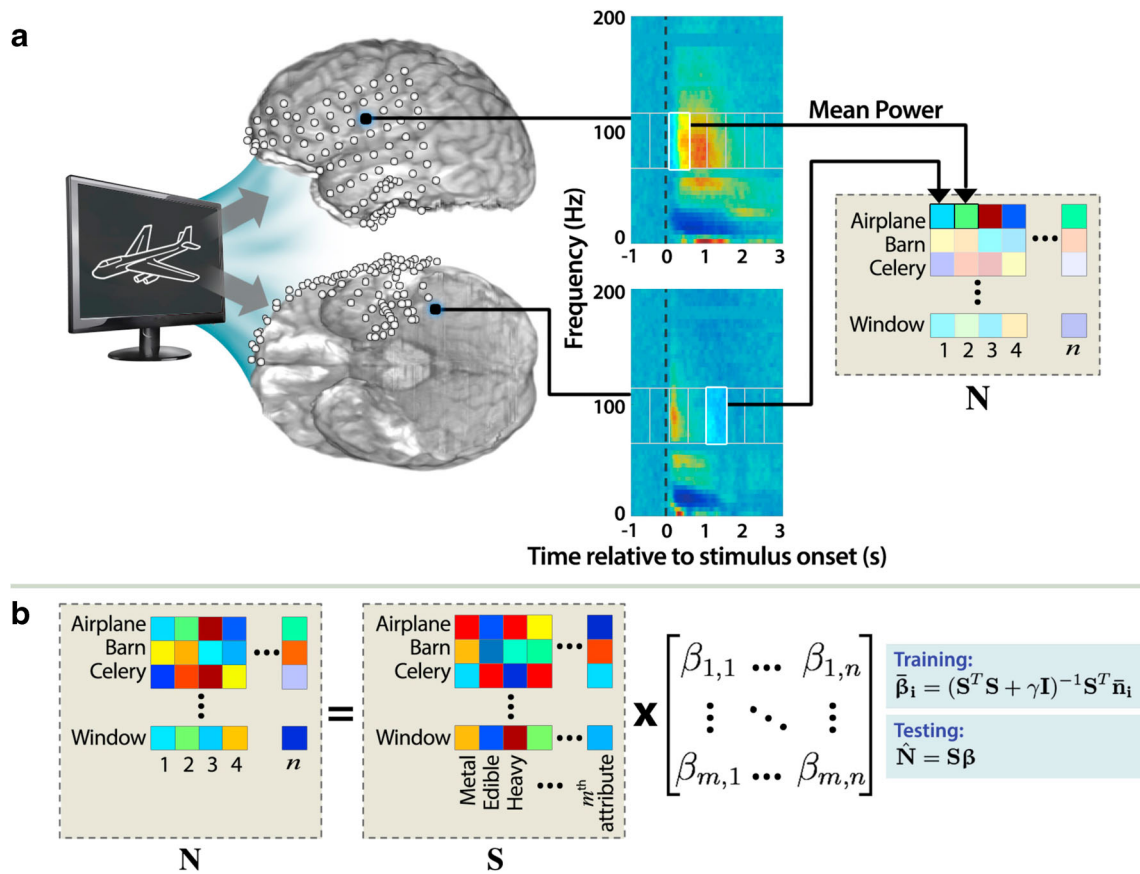


Fig. 3 Semantic feature-encoding model used by Rupp et al. [21]. (a) Mean spectral power (numbered 1 to n) during particular temporal windows in specific frequency bands in specific channels collected while a subject views and names images from each semantic category can be concatenated into a neural feature vector (N) for each semantic category [21]. (b) Reverse-decoding scheme, in which expected neural activity is predicted from a semantic feature-encoding model [21]. Semantic attributes (numbered 1 to m) corresponding to each semantic category can be concatenated into a semantic feature vector (S), which can be used in a

neural encoding model (β), in this case through linear ridge regression, to predict N from S [21]. In this reverse-decoding process, which can often be more effective, new neural activity patterns are compared to the characteristic “template” patterns generated by β and assigned to the closest category [21]. Note, however, that it would also be valid to predict semantic features from neural activity and compare them to the characteristic semantic feature vectors instead. Figure reused with permission from Ref. [21]

study, object-category information was decoded in single trials in ECoG using activity in the temporal and occipital lobe, suggesting the presence of visual-semantic integration in these areas tied to semantic category recognition. Follow-up research by Wang et al. [34] and Manning et al. [35] confirmed this result respectively by predicting semantic category in single trials using only speech-production-locked activity in the high gamma band (60–120 Hz) in the left inferior frontal gyrus (IFG) and posterior STG and by finding that semantic similarity of neural activity exhibits a distinct band-diagonal structure, implying near-optimal self-similarity, and that it is highly correlated with the semantic similarity of the words themselves just prior to word recall, likely reflecting speech intention (see Fig. 6). The results of Wang et al. [34] led them to suggest the potential for a semantics-based speech BCI, the first known reference to such a concept in the literature.

The previous ECoG studies were then, after a short gap in the literature, followed by two more recent studies [21, 36]. Chen et al. [36] observed that stimulus-locked activity in the ventral temporal lobe encodes semantic representations uncorrelated with the visual similarity of the images presented or phonological similarity of the names of the objects shown. Rupp et al. [21] found that semantic-attribute-encoding models can be used to decode untrained objects with ECoG using high gamma activity (70–110 Hz) in the basal occipitotemporal area along specific semantic dimensions with accuracies comparable to whole-brain fMRI. Further, they found that individual patient results closely agree with reports from other imaging modalities on the time course and functional organization of semantic processing along the ventral visual pathway during object recognition [21]. These findings are consistent with previous research defining the ventral temporal lobe as a key area encoding

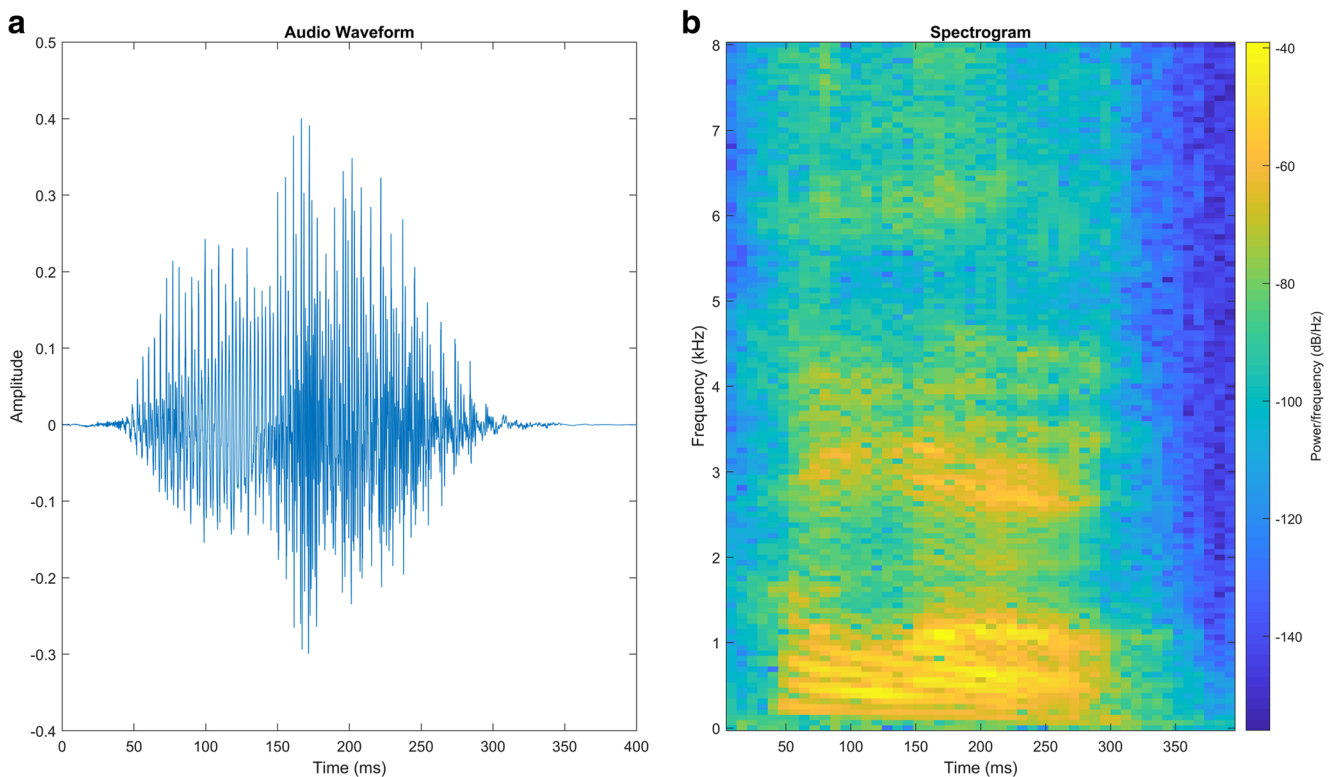


Fig. 4 Spectrotemporal representations of speech acoustics. (a) Raw audio waveform of “hello” uttered by a male speaker. (b) Spectrogram of “hello” uttered by a male speaker. Changes in the audio signal for each phoneme are reflected in both the temporal (a) and spectrotemporal (b) domains

visual object semantics but suggest that further research is necessary to determine how semantic dimensions are encoded anatomically within the brain. Further, they suggest that semantics can be decoded with high accuracy in single trials and thus potentially in real time.

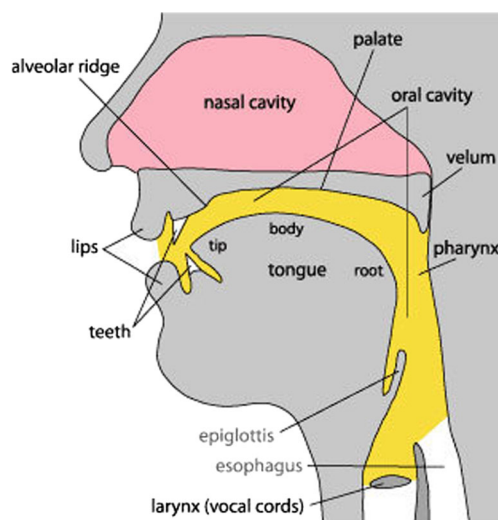


Fig. 5 Key speech articulators in the human vocal tract. In the source-filter model of speech production, the vocal cords (the source) vibrate to generate sound, which is then modulated by the positions of the speech articulators (the filter) over time to generate specific speech sounds (i.e., phonemes) [25]. These speech articulators and the vocal cords are shown above. Figure reused with permission from Ref. [26]

Auditory Decoding

Neuroimaging research has contributed to an increasingly detailed model of the cortical regions in and around the STG responsible for speech perception, including those that could be of utility for speech BCI [37]. Recent studies using fMRI, EEG, and MEG have found evidence for three tonotopic gradients in the auditory cortex, two stronger tonotopic gradients in the rostral and caudal Heschl’s gyrus, which includes the primary auditory cortex (A1), forming a “V” shape, and one weaker gradient in the planum temporale within Wernicke’s area, as illustrated in Fig. 7 [38–40]. However, other studies suggest that neural responses in the auditory cortex can reflect perceived pitch, which differs based on harmonics present in the signal, in addition to raw frequency components of acoustic stimuli [41, 42]. Thus, a pitch-adjusted spectral target for an auditory speech BCI may align better with components of the natural auditory response and provide maximal performance.

Several ECoG studies suggest a reliable correlate of the finely graded auditory response is captured by the coarser population-level activity resolved by ECoG [43–45]. Mesgarani et al. [43] and Hullet et al. [44] observed that the STG exhibits a distributed population response in the high gamma band (70–150 Hz) to continuously perceived speech with a robust anterior–posterior spatial distribution of spectrotemporal tuning encoding acoustic cues in a nonlinear fashion

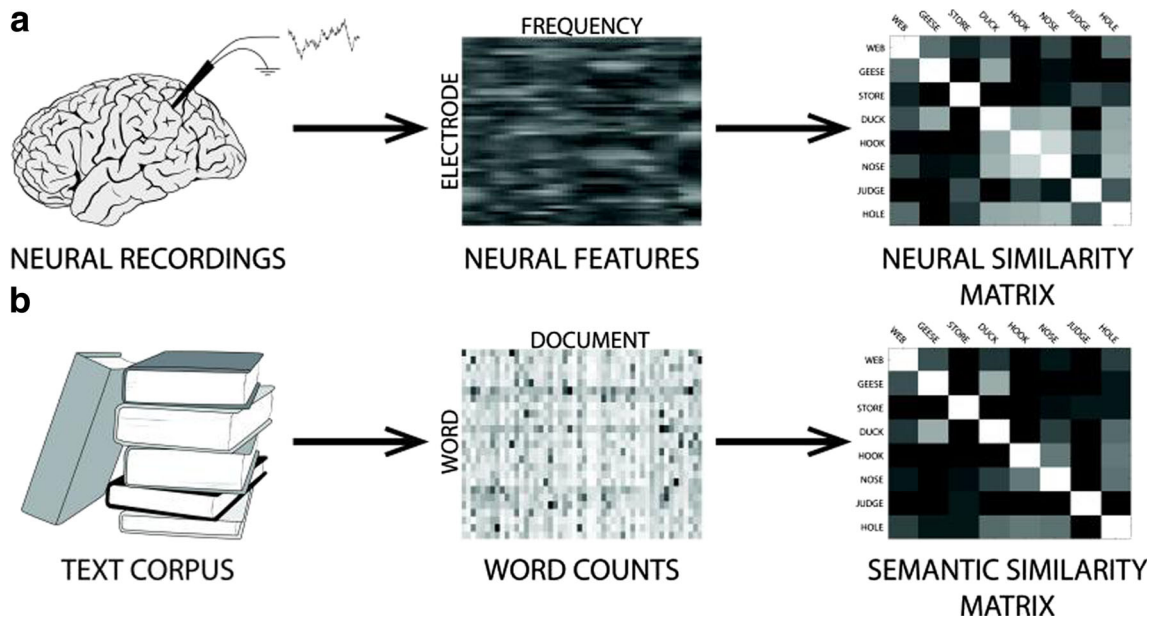


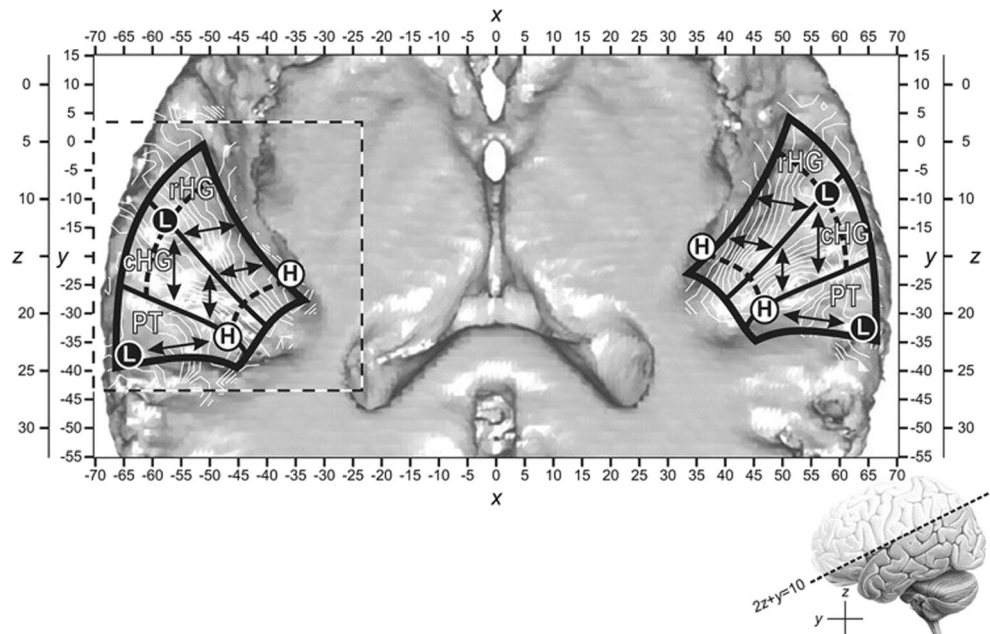
Fig. 6 Semantic organization in cortex according to Manning et al. [35]. (a) Mean power at specific frequencies are calculated from neural recordings during word presentation and recall and then passed through principal components analysis (PCA) [35]. The PCA components that vary systematically with the meanings of the presented words are used to construct a neural similarity matrix by measuring the pair-wise cosine similarities between the PCA feature vectors for each word [35]. (b) Semantic features of the words themselves are calculated through latent

semantic analysis (LSA) [35]. A semantic similarity matrix is constructed by measuring the pair-wise cosine similarities between the LSA feature vectors for each word [35]. Note that the neural similarity (a) displays a distinct band-diagonal structure, implying near-optimal self-similarity [35]. This, along with the finding that neural and semantic similarity are highly correlated just prior to word recall, suggests that words are encoded in cortex based on semantic features during speech intention [35]. Figure reused with permission from Ref. [35]

or through a combination of cues. In particular, the posterior STG is more highly tuned for temporally fast varying speech sounds with relatively constant energy across the frequency axis (i.e., low spectral modulation), whereas the anterior STG is more highly tuned for temporally slow varying speech sounds with a high degree of spectral variation across the

frequency axis (i.e., high spectral modulation), as illustrated in Fig. 8 [44]. On the other hand, Hamilton et al. [45] observed that spatially localized, caudal, or posterior regions of the STG are more sensitive to speech onset, while spatially distributed, rostral or anterior and medial regions of the STG are more sensitive to sustained speech. Together, these studies suggest

Fig. 7 Tonotopic organization in the auditory cortex. Two, stronger, tonotopic gradients from low- (L) to high-frequency (H) encoding auditory information are present in the rostral (rHG) and caudal Heschl’s gyrus (cHG), forming a “V” shape, and a third, weaker, tonotopic gradient is present in the planum temporale (PT) [38, 39]. The above slice is shown from a top-down view of the superior-temporal plane depicted in the bottom-right corner. Figure reused with permission from Ref. [39]



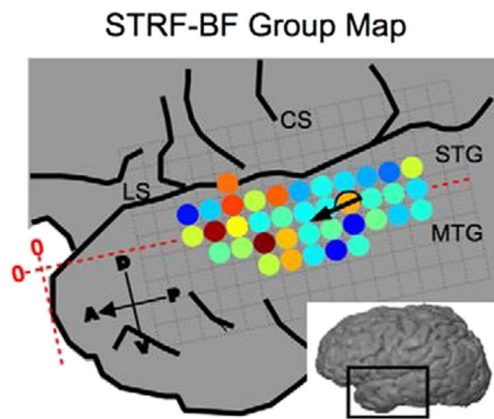


Fig. 8 Spectral tuning along the superior temporal gyrus according to Hullet et al. [44]. Above is a map of the maximally activated “best frequencies” (BF) of the spectrotemporal receptive fields (STRFs) of subpopulations of neurons within the superior temporal gyrus (STG) [44]. Spectral tuning runs along the STG in the anteroventral direction from high (1000 Hz, red) to low (125 Hz, blue) spectral modulation [44]. The STG is surrounded by the lateral sulcus (LS), central sulcus (CS) and middle temporal gyrus (MTG) [44]. Figure reused with permission from Ref. [44]

that ECoG adequately captures spectrotemporal tuning of the STG to speech organized by acoustic features rather than by discrete phoneme categories.

Silent reading tasks in fMRI and ECoG suggest that covert speech shares a substrate with perceived speech [46, 47]. Specifically, these studies observed 1) that activity during silent reading is significantly correlated with BOLD [46] and high gamma activity (50–150 Hz) [47] in temporal voice-selective areas (TVAs), such as the A1, posterior and medial superior temporal sulcus (STS), and ventral occipital temporal cortex [46, 47], and 2) that there is considerable jitter between activation peaks in single trials in visual and auditory cortices, suggesting that covert reading is both a visual and auditory process with each process occurring simultaneously but independently [47]. This activity is greater for direct *versus* indirect passages (i.e., first- *versus* third-person descriptions) and positively modulated by subject attention [46]. Further research suggests a fine grapheme–phoneme correspondence (GPC, i.e., sound–letter mapping) [48, 49]. In an ECoG study by Perrone-Bertolotti et al. [48], the STG and IFG were implicated in simple and complex phoneme detection modulated by GPC. Similarly, in an MEA study by Chan et al. [49], firing of several STG units to visual letters correlated with specific auditory phonemes (a la GPC). Together, these findings suggest that covert speech can be decoded phonemically. However, they also suggest that optimally training and testing a decoder based on auditory representations of speech production may require the subject to evoke the auditory experience that accompanies speech production, which may be unnatural and effortful for some.

Several ECoG studies explicitly study the relationship between overt and covert speech [50–52]. Magrassi et al. [50]

observed that activity in the theta band (2–8 Hz) was maximally correlated to the speech envelope (likely reflecting prosodic or syllabic rate) prior to articulation in both overt and covert reading, whereas activity in the high gamma band (range not given) was also highly correlated to the speech envelope (likely reflecting phonetic rate and words in isolation) but less so than activity in the theta band. Similarly, Brumberg et al. [51] observed that activity in the high gamma band (70–170 Hz) progresses from speech motor areas in the ventral precentral gyrus and Broca’s area to auditory speech areas in the middle temporal gyrus (MTG) and middle and posterior STG (including the A1 and Wernicke’s area) during both overt and covert reading. Additionally, Martin et al. [52] observed that a model built to reconstruct auditory speech features from high gamma activity (70–150 Hz) during overt speech could reconstruct these same features during covert speech. The common areas activated during overt and covert speech are shown in Fig. 9, and the speech reconstruction results are shown in Fig. 10. These results suggest a shared auditory and articulatory substrate between overt and covert speech with a spatiotemporal progression of activity aligned to speech production.

However, further studies suggest a differential contribution of auditory and articulatory areas to overt and covert speech production over time [53–55]. In an MEG study by Tian and Poeppel [55], it was observed that overt and covert speech articulation are fundamentally different but that overt and covert perception are highly similar. Specifically, they observed that activation during covert articulation extends to posterior parietal areas rather than the motor cortex, as in overt articulation, but that auditory activation in bilateral temporal areas is present during both overt and covert speech perception [55].

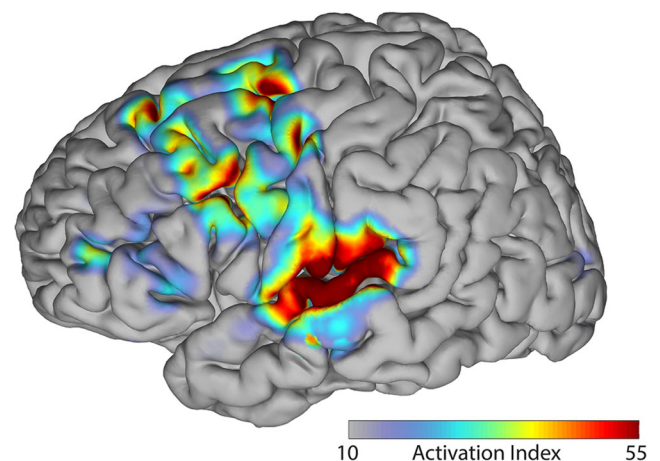
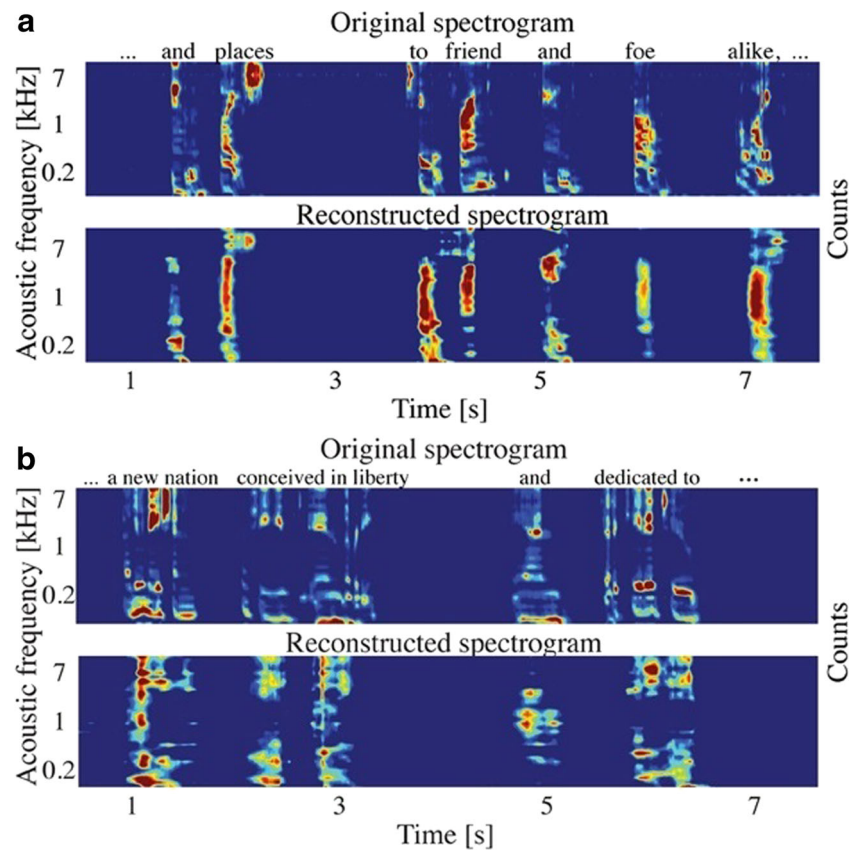


Fig. 9 Shared neural substrate between overt and covert speech. Shared activation between overt and covert speech is located primarily in the superior temporal gyrus, premotor cortex, and primary motor cortex [51]. Shared activation can also be observed in the frontal eye fields and dorsolateral prefrontal cortex; however, this may be indicative of visual processing and decision making [51]. Figure reused with permission from Ref. [51]

Fig. 10 Auditory spectrogram reconstruction. (a) Original and reconstructed spectrograms for overt speech [52]. (b) Original and reconstructed spectrograms for covert speech [52]. Because there is no ground-truth audio signal for covert speech, the spectrogram of the overt speech that would have been spoken is used in place of the original [52]. Figures reused with permission from Ref. [52]



Covert speech articulation also displays an “auditory-like” response following articulation highly similar to that observed during hearing, suggesting the presence of an “efference copy,” a feed-forward prediction of the perceptual outcome of an action [55]. These findings are supported in ECoG studies by Pei et al. [53] and Leuthardt et al. [54], in which both overt and covert speech repetition tasks using both auditory and visual cues (4 combinations) were performed. In these studies, areas within the STG, including the planum temporale within the A1 (BA41/42) and Wernicke’s area (BA22), showed more pronounced activation during covert speech than in the primary motor cortex, with some disagreement about the premotor cortex [53, 54]. However, the disagreement about the premotor cortex was cleared in ECoG studies by Pei et al. [56] and Ikeda et al. [57] studying phonemes in isolation, in which high gamma activity (70–110 Hz) in the premotor cortex and the STG contributed most to decoding performance for covert speech. Of particular note is that unlike previous studies, Ikeda et al. [57] did not find significant activation in Broca’s area, which led them to postulate an involvement of Broca’s area in sequencing phonemes [57]. Together, these studies suggest that a tightly controlled covert speech paradigm is necessary to ensure consistent results and that auditory approaches may perhaps be more readily adaptable to covert neural speech decoding. Additionally, they suggest that decoding phonemes in context may differ

from decoding them in isolation. However, further research is necessary to identify the precise contributions of the different types of activity to decoding performance over time and the potential integratory role of Broca’s area in sequencing phonemes.

Articulatory Decoding

It is well established in the literature that the muscles of the human body are somatotopically mapped in the motor cortex [58, 59]. This mapping has been utilized in previous BCI studies of upper-limb movements [60–62] and naturally extends to speech articulators, as evidenced by multiple ECoG studies [63–65]. Bouchard et al. [63] found that temporally locked activity in the high gamma band (85–175 Hz) in the ventral sensorimotor cortex (vSMC) is organized spatially by articulatory features, following a dorsal–ventral layout mimicking the rostral-to-caudal layout of the vocal tract during overt reading, and that a coordination of multiple articulator representations across the vSMC network is necessary to generate speech (see Fig. 11). Toyoda et al. [64] found that variably timed high gamma augmentation and attenuation (70–110 Hz) at distinct sites corresponded to particular phonemes during overt isolated-phoneme articulation and that voice-onset-locked activity in larynx-sensorimotor areas differs between voiced and unvoiced phonemes. Lotte et al. [65] found

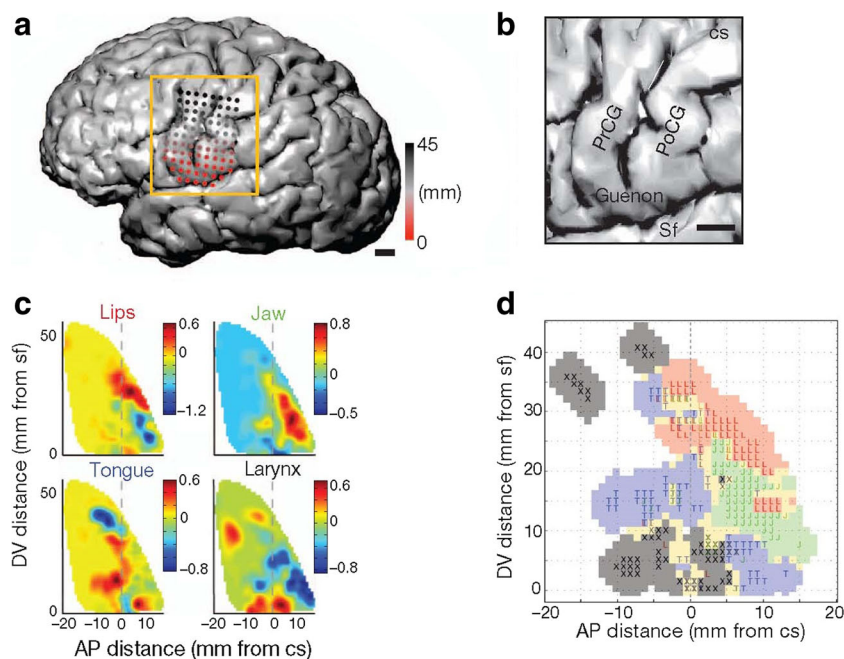


Fig. 11 Somatotopic organization in the ventral sensorimotor cortex according to Bouchard et al. [63]. (a) Electrodes over the ventral sensorimotor cortex (vSMC) are colored based on their distances from the Sylvian fissure (Sf) [63]. (b) A close-up view of the vSMC in relation to the precentral gyrus (PrCG), postcentral gyrus (PoCG), central sulcus (CS), Sf, and subcentral gyrus (as corrected from guenon in the erratum) [63]. (c) Localization of speech-articulator representations, where the

color scale corresponds to the average magnitude of the articulator weightings along the dorsal–ventral (d–v) and anterior–posterior (a–p) axes [63]. (d) Functional somatotopic organization of the speech articulators with the following color correspondence: lips (L, red); jaw (J, green); tongue (T, blue); larynx (X, black); mixed (gold) [63]. Figures reused with permission from Ref. [63]

that phonemes can be decoded during continuous overt speech using high gamma activity (70–170 Hz) in speech sensorimotor areas using articulatory correlates, such as place of articulation, manner of articulation, voicing, and vowel-*versus*-consonant classification. Together, these studies suggest that temporally locked, somatotopically mapped activity in the vSMC can distinguish phonetic segments of speech based on articulatory features, which suggests that this approach may be highly informative for real-time neural speech decoding.

Several studies investigate to what extent activity in the vSMC is correlated with speech articulation *versus* speech perception [66–68]. Cheung et al. [66] found in ECoG that, in contrast to the somatotopic map during overt speech production, an acoustically based topographic map similar to that in the STG emerges during speech perception in the high gamma band (70–150 Hz). The negative finding is backed by Arsenault and Buchsbaum [67], who found in fMRI that univariate and multivariate analyses failed to find evidence for somatotopic coding in the motor or premotor cortex, which partly include the vSMC, during speech perception when compared to mimed speech, a paradigm which isolates articulation from the confound of self-perception. On the other hand, Glanz Iljina et al. [68] found that high gamma activity (70–350 Hz) was present in an electrostimulation-defined mouth motor region in the superior ventral premotor cortex,

which partly includes the vSMC, both during speech production early relative to speech onset and during speech perception. Further, they found that this activity was not present during nonspeech perception and that it was not significantly modulated by acoustic background noise [68]. Together, these results suggest a role of the vSMC during speech perception (and thus potentially covert speech) possibly related to the envisioning of speech intention. However, further research is necessary in this area.

Multiple studies have shown that an articulatory approach to neural speech decoding is feasible for BCI [69–71]. Tankus et al. [69] found using hybrid (clinical + microwire) depth electrodes that single neurons in the rostral anterior cingulate and adjacent medial orbitofrontal cortex are sharply and specifically tuned to individual vowels and that the firing of neural populations to vowels is organized according to their place and manner of articulation as reflected by the International Phonetic Alphabet (IPA) chart (as inferred from the error structure of their neuronal classifier). The latter result is illustrated in Fig. 12. Because the IPA chart corresponds to the position of the highest point of the tongue during articulation, this organization of speech features corresponds to a functional spatial–anatomical organization according to tongue height [69]. This finding is supported by an ECoG study by Mugler et al. [70], which found that overtly spoken consonants are

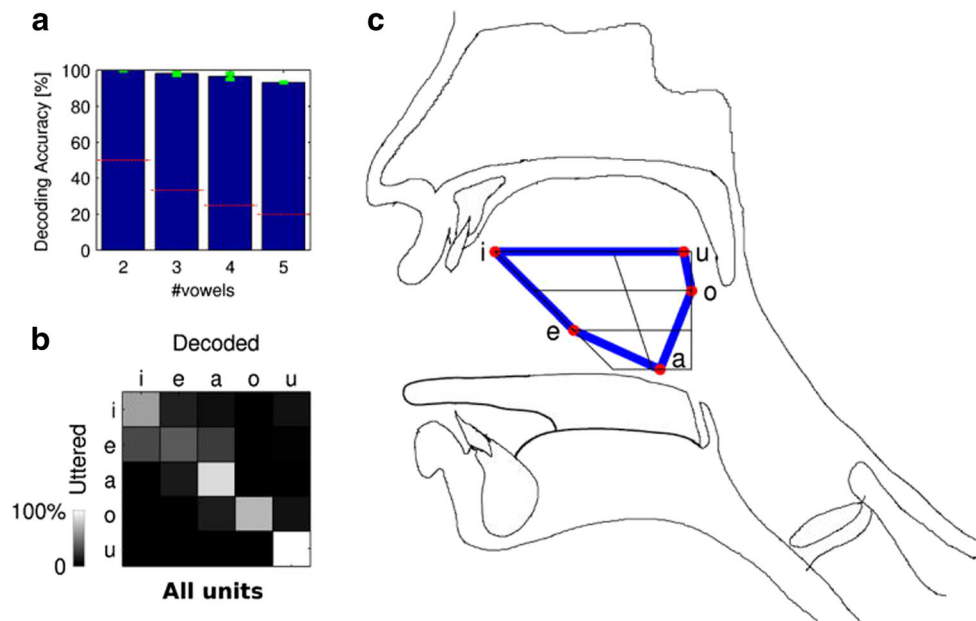


Fig. 12 Concordance of neural activity organization with tongue height as observed by Tankus et al. [69]. (a) Decoding accuracy of vowels with standard error bars shown in green and chance-level decoding accuracy shown as dashed red lines [69]. (b) Confusion matrix for vowel decoding illustrating pair-wise confusability when decoding vowels [69]. (c) Tongue position within the vocal tract for each vowel according to the International Phonetic Alphabet (IPA) chart [69]. The vowel confusion

matrix (a) displays a distinct band-diagonal structure with maximal error between neighboring vowels when ordered based on the place of the tongue during articulation in accordance with the IPA chart (b), which suggests the presence of a functional spatial–anatomical organization of neural activity according to tongue height [69]. Figure reused with permission from Ref. [69]

organized according to their place and manner of articulation as reflected by the IPA chart (as inferred from the error in the confusion matrices from their decoding results). The best-performing electrodes were found in speech motor areas along the posterior frontal lobe around the vSMC, with performance relying on precise temporal alignment to phoneme onset and use of activity in the high gamma band (65–250 Hz) [70]. Further, Mugler et al. [70] obtained 36% classification accuracy for phonemes overall and up to 63% for single phonemes, allowing word identification equivalent to information transfer rates of up to 3.0 bits/s (33.6 words/min), which they deemed sufficient to support the use of speech articulation for BCI control. In a follow-up ECoG study, Mugler et al. [71] found that, despite their inherent correlation, articulatory gestures could be decoded using high gamma activity (70–300 Hz) in speech motor areas with significantly higher performance than phonemes at all electrodes, particularly in more posterior areas of cortex, rather than more anterior areas in which performance was similar. Together, these findings suggest the potential for an articulatory speech BCI and the potential superiority of articulatory-gesture features when decoding from speech motor areas.

A variety of ECoG studies support the assertion that articulatory representations in the vSMC contain sufficient information for high-performance speech BCI. For example, Bouchard and Chang [72] accurately predicted 81% of the acoustic variability across vowels during overt syllable production in single

trials using activity primarily from the high gamma band (70–150 Hz) in the vSMC. Similarly, Kanas et al. [73] achieved 98.8% accuracy with the same experimental paradigm in voice activity detection (VAD) using a support vector machine (SVM) classifier, with high gamma being the most informative band (88–144 Hz) and the posterior STG, STS, and vSMC containing the most informative electrodes. At the word level, Mugler et al. [74] decoded phoneme position in word context during overt word production based on differences in peak high gamma timing (70–300 Hz) with 76.94% accuracy. At the sentence level, Zhang et al. [75] correctly classified 77.5% of trials during overt sentence repetition using dynamic time warping in combination with an SVM classifier and high gamma activity (60–90 Hz) primarily in the posterior IFG. These results suggest that relatively high performance can be achieved with relatively small training times, but do not elucidate to what extent articulatory information alone contributed to the performance, requiring further research.

Feasibility of Speech BCI

Overt neural speech decoding, though difficult, has shown the most progress to date and may be most applicable for patients with total LIS for whom intercepting overt speech (effectively bypassing the neural pathways to the speech articulators) would be adequate. Overt and covert neural speech decoding has typically been done using neural activity in the high

gamma band, which has consistently been shown to be the most informative band for neural speech decoding. The exact limits of this band are unknown, with previous studies successfully utilizing frequencies as low as 50 Hz and as high as ~100, 200, or even 300 Hz. Nonetheless, this finding directly aligns with previous research studies, which suggest that high-frequency activity corresponds most with the firing of individual neurons within a population [76–78]. On the other hand, covert neural speech decoding, which may be necessary for patients with less-severe cases of LIS, has proven extremely challenging, largely due to weaker activation observed during covert *versus* perceived and overt speech and the lack of a clear response to align to as with overt speech, e.g., acoustic or articulatory recordings. For this reason, the refinement of covert neural speech decoding will depend largely on novel experimental paradigms and the translation of techniques from overt neural speech decoding. For more in-depth discussion, see Martin et al. [79].

Two chronic studies have been performed to date, in which subjects with LIS have been trained to control a BCI facilitating communication, including one by Vansteensel et al. [80] using motor-BCI-based communication system with a limited number of electrodes to type individual letters and one by Guenther et al. [81] to control a real-time speech synthesizer for select vowels using a single neurotrophic electrode. Of particular note is that Guenther et al. [81] was able to increase the subject's accuracy by 25% (from 45 to 70%) and decrease his average endpoint error in a block paradigm by 46% using auditory feedback from the decoded sound. This suggests that feedback training may be key to successful speech-BCI training.

No chronic neural speech decoding studies using ECoG have been performed to date; however, multiple acute studies have demonstrated the potential feasibility of such an approach. Herff et al.'s [82] "Brain-to-Text" system was the first demonstration of a working system, which transcribed overt speech intention inferred from widely distributed brain areas directly into text with word and phone error rates ranging 25 to 60% and 50 to 80% for vocabularies ranging 10 to 100 words respectively (all significantly above chance). Though their performance was relatively modest, it serves as a useful benchmark to understand how performance might scale with increasing vocabularies. Additionally, follow-up analysis demonstrates that these above-chance results extend to decoding using only prephonatory activity, a rough analog for decoding covert speech intention according to the authors [83]. This was followed by two studies demonstrating with Moses et al. [84] and Moses et al.'s [85] NSR ("Neural Speech Recognition") system 1) that utilizing both spatial and temporal information and incorporating language modeling and sequential path probabilities improves neural speech decoding performance and 2) that sentences can be classified with 90% accuracy using a unique online block-wise

retraining and classification scheme with limited vocabularies of 10 sentences (both during speech perception). Though it is unclear how well these results will generalize in pretrained models, high-performance decoding in real time may be possible given further refinement of neural speech decoding approach.

A variety of approaches for spectral feature extraction and classification have been used for speech BCI thus far. The typical scheme involves first extracting spectral energies in specific frequency bands from each channel through a fast Fourier transform (FFT), bandpass filter bank, or wavelet transform. Though each method possesses its own design tradeoffs due to its formulation, the FFT, bandpass filter, and wavelet approaches are ultimately mathematically equivalent [86]. These extracted features are then either regressed (or in other words mapped) to semantic, auditory, or articulatory parameterizations of the target speech or used to classify utterances on a frame-by-frame or entire-trial basis. Regression and classification can be done with a variety of algorithms, from those as simple as support vector machines or linear discriminant analysis to those as complex as hidden Markov models or deep neural networks. A comprehensive review of these machine learning approaches is outside of the scope of this paper; however, all such algorithms ultimately accomplish the same goal, that is to learn how to accurately and consistently convert some parameterization of the subject's neural activity into a parameterization of the corresponding speech signal, whether that consists of features that can be fed into a speech synthesizer or simply labels describing the words that comprise that speech. We encourage the inquisitive reader to read through the methods of the cited papers in our neural speech decoding sections.

Leveraging multiple levels of speech and language representation may further increase the performance of speech BCIs. This approach is outlined in a proof-of-concept study by Bocquelet et al. [87], in which subjects were trained to control a speech synthesizer with overt and semi-overt speech paradigms while undergoing electromagnetic articulography (EMA) recording, which was fed into a deep neural network to map articulations to their corresponding acoustic outputs. This study demonstrated intelligible, real-time synthesis of vowels and consonants by compensating for differences in sensor positions and articulatory differences between new speakers and the reference speaker [87]. Although their methodology was not tested with neural data, the authors suggested that anatomical differences could be accounted for without extensive retraining and that parameters derived from EMA activity could be inferred from recorded neural activity [87]. Their proposed methodology is outlined in Fig. 13. The feasibility of this approach with ECoG was investigated by Bouchard et al. [88] and Conant et al. [89], who decoded vocal tract kinematics during overt speech from the vSMC using activity in the high gamma band (70–150 Hz), and Herff

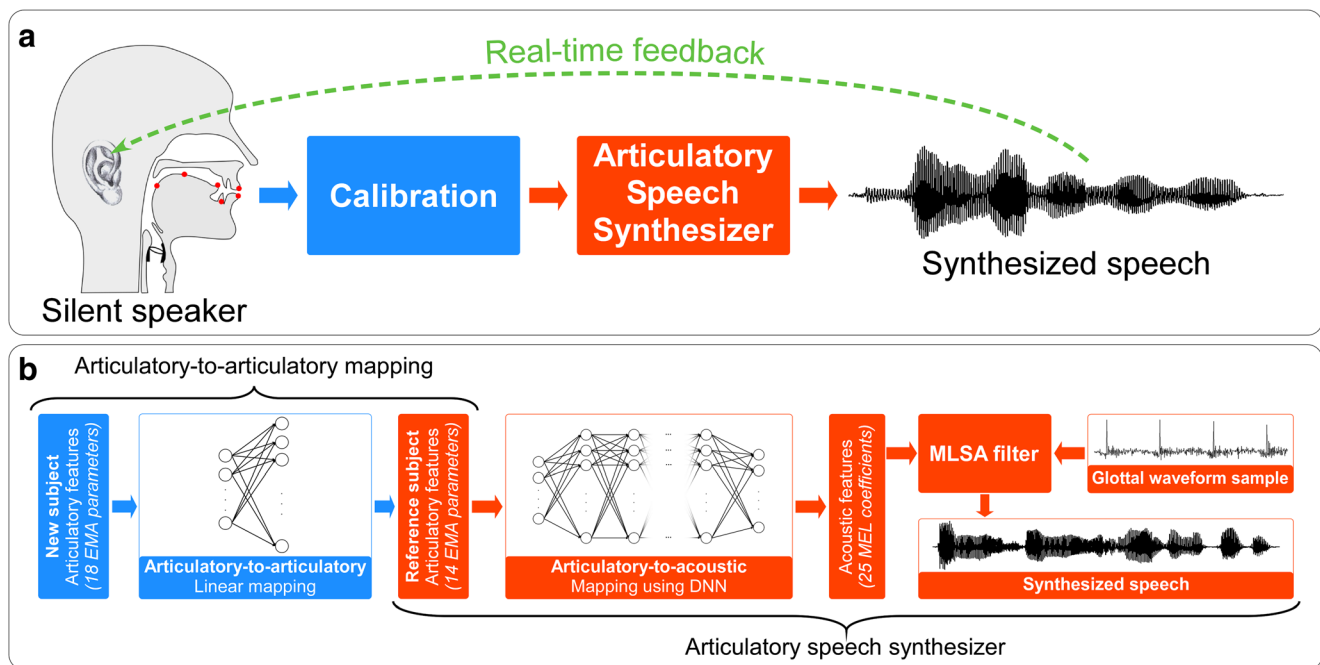


Fig. 13 Strategy proposed by Bocquelet et al. [87] to synthesize speech from articulatory representations in the brain. (A) Overall flow of training approach, in which a silent speaker synthesizes speech with an articulatory speech synthesizer calibrated to the anatomy of the speaker and based on a real-time feedback loop. (B) Specific steps of articulatory-to-acoustic inversion approach, in which articulatory features are adjusted to match a reference speaker and then passed through a deep neural network to obtain acoustic features that can be excited by an Mel Log Spectrum

Approximation (MLSA) filter to synthesize speech [87]. In the original study, articulatory parameters derived from electromagnetic articulography (EMA) recordings were mapped to the intended speech waveform [87]. However, the authors suggest that in a BCI application, the EMA parameters could be inferred from neural activity and then mapped to the intended speech waveform using the pretrained ground-truth model [87]. Figure reused with permission from Ref. [87]

et al. [90], who decoded speech spectrograms during overt speech from around the STG also using activity in the high gamma band (70–170 Hz). Both demonstrated that decoded activity can be used to resynthesize speech with high correlation to the original speech signal, with the former showing that vowels can be decoded with higher accuracy using anatomical information than can be done using traditional methods [88–90]. However, the most successful demonstrations to date are arguably by Akbari et al. [91] and Milsap et al. [92]. Akbari et al. [91] demonstrated that highly intelligible speech can be resynthesized from the A1; the best performance was achieved using deep learning, both low- and high-frequency information and a vocoder target, which provided 75% intelligibility and 65% relative improvement over the linear-classifier, spectrogram-target baseline. On the other hand, Milsap et al. [92] demonstrated that 100% neural VAD of isolated syllables, with decoding performance likely limited only by electrode placement and density of coverage and less than ~1-s latency, is achievable through a simple matched-filter approach. Further, they observed that activity in the vSMC best discriminates place of articulation and consonant voicing, whereas activity in the STG best discriminates vowel height [92]. Together, these results suggest that high-performance speech BCI with high enough performance to justify a chronic ECoG study is possible. Further, they suggest that an approach

that integrates multiple levels of speech and language representation may allow for increased performance over simpler, agnostic, classification methods.

Although the discussed results show promising algorithmic approaches for speech BCI, there remain significant challenges that must be addressed. A truly functional speech BCI will need to rely primarily on speech-production-or intention-based cues. Truly continuous neural speech decoding remains elusive, and most results with speech production have been in decoding very restricted vocabularies of phonemes, syllables, words, or sentences. Though highly intelligible speech can be synthesized during speech perception [91], speech generated from cortical articulatory representations has thus far only shown to be correlated to the ground-truth speech signal but not intelligible [88–90]. Though covert speech shares major neural representations with perceived speech, high-quality, covert neural speech synthesis using algorithms trained on perceived speech has not been demonstrated. Significant algorithmic breakthroughs will be necessary before continuous neural speech decoding can allow a patient to converse freely. However, continuous neural speech decoding is not necessary for speech BCI to benefit patients. Detecting isolated keywords or short phrases from ECoG activity is a much more tractable problem and may be sufficient to allow patients to control their devices or communicate basic

needs and desires through simple neural voice commands. Although a recent study of this by Milsap et al. [92] was limited to detecting syllables in isolation rather than full keywords in naturalistic settings, good neural VAD and discriminability suggests the promise of this approach. Finally, the results obtained thus far have been achieved with limited data collected acutely from patients with suboptimal coverage of cortical areas important for speech production. It is likely that extended or long-term studies in individual patients will be needed to fully explore the potential capabilities of ECoG-based speech BCI.

Safety and Efficacy of Chronic ECoG

Chronic ECoG in Animals

Studies in both rodents and primates have shown that subdural ECoG arrays with dozens of electrodes can be used to record and decode neural signals with acceptable frequency characteristics reliably over several months [93–95]. Specifically, both arm motion (often used as the basis for motor BCIs) and computer-mouse control can be decoded reliably with no drift in accuracy or performance drop-off or need for recalibration [94, 95]. The long-term stability of ECoG signal quality and decoding was also confirmed in the most comprehensive study to date by Degenhart et al. [96]. In this study, a single monkey was implanted with a high-density subdural ECoG grid over cortical motor areas for 666 days to determine whether ECoG signals can be recorded reliably long-term without significantly damaging underlying cortical tissue. Histological analysis revealed minimal damage to cortex beneath the implant, though the grid was encapsulated in collagenous tissue, and macrophages and foreign body giant cells were observed at the tissue-array interface [96]. Further, despite evident encapsulation and foreign body response, the authors obtained stable cortical recordings for more than 18 months following implantation. Thus, in combination with previous studies, this study suggested not only that ECoG recordings can remain stable over a period of years but also that host tissue response does not significantly hinder signal acquisition [96].

Long-Term ECoG-Based Neurostimulators in Humans

Though past human ECoG–BCI studies have not lasted for longer than a month, ECoG–neurostimulator studies can serve as useful reference for signal stability and complication rates with chronic ECoG. The longest study with the most electrodes was that of Duckrow and Tchong [97], in which 24 human subjects with epilepsy were implanted chronically with a combination of subdural ECoG strips, grids, and depth electrodes, consisting of anywhere from 24 to 208 electrodes, as well as a NeuroPace-RNS neurostimulator system, and monitored

anywhere from 2 to 54 days to observe the daily variation in their ECoG features. The authors successfully recorded activity variation across brain regions by varying the active electrodes and did not note any side effects beyond postoperative pain [97]. This study suggests that extensive ECoG arrays with hundreds of electrodes could potentially be implanted safely for up to at least a couple months in humans.

The viability for ECoG has been demonstrated for longer time spans as well, albeit with a much smaller number of electrodes. A variety of studies have been performed retrospectively in up to 191 patients implanted with NeuroPace-RNS or NeuroVista neurostimulator systems with 8 or 12 electrodes for as long as ~7 years [98–100]. All studies found that activity remained stable over their period of implantation; however, only Cook et al. [98] commented on clinical effectiveness or rate of complication [98–100]. In Cook et al.'s [98] study, no changes in clinical effectiveness were observed 4 months after implantation with only 3 significant complications [98]. These included a device migration and seroma immediately following implantation, with the device repaired uneventfully; an infection at 7 months, leading to explantation due to lead damage; and discomfort due to lead tautness, leading to explantation at the patient's request [98]. Based on these studies, it is possible that a subject could be safely implanted with ECoG for 6 months to a year or more, long enough to establish the safety and efficacy of a chronic ECoG-based speech BCI with a greater electrode count.

Rao et al. [101] recently demonstrated that ECoG signals recorded chronically with an RNS system can be used for neural speech decoding. Specifically, they observed that the signals remained stable over 1.5 years and that they contained useful speech correlates, including selectivity to phonetic features of spoken sentences in Wernicke's area and cortical evoked responses in Broca's area preceding speech onset during overt word naming [101]. This finding suggests that an ECoG-based speech–BCI system may continue to perform well even over a period of years.

Epidural Versus Subdural ECoG

The relative utility of epidural and subdural ECoG has been debated within the ECoG literature. Although the majority of studies mentioned in this review used subdural grids, studies, such as that Rouse et al. [95], have demonstrated that decoding is possible in monkeys using epidural implants. Epidural implants are somewhat less invasive than subdural implants, because they do not require breaching the dura, reducing the likelihood of rare complications such as meningitis, cerebritis, and venous infarction, as well as more common foreign body tissue reactions [95]. However, the dura mater may significantly attenuate neural signals and decrease decoding performance [102]. Bundy et al. [102] observed that micro-scale, subdural electrodes have significantly higher

spectral amplitudes and reach the noise floor at a higher frequency than micro-scale, epidural electrodes do, but that for macro-scale electrodes, the statistical differences are small in amplitude and likely irrelevant for BCI systems. Flint et al. [103] observed that epidural decoding of hand grasps is comparable to subdural decoding. Together, these studies suggest that larger, more sparsely spaced, epidural electrodes are comparable to larger, more sparsely spaced, subdural electrodes. However, this may not be the case for smaller, more densely spaced electrodes, which have shown useful for speech–BCI applications in humans [104] and for which further research is necessary.

Risks of Acute ECoG Implantation

ECoG implantation for seizure monitoring prior to epilepsy surgery can serve as a useful reference point for the safety of implanting more electrodes than those of the RNS and NeuroVista systems mentioned, which typically implant no more than 8 to 12 electrodes. Multiple studies have assessed complication rate in hundreds of patients from acute intracranial monitoring, which typically employs 32 to 128 or more electrodes, and its change from past to present [105–110]. In their study, Hamer et al. [106] noted that overall complication rate from 1980 to 1997 decreased from 26.3% over the past 23 years, including 1 mortality after grid implantation, to 19% over the past 5 years, including 2% permanent complications, and 13.5% over the past 3 years, with no mortalities or permanent complications. Similarly, Nagahama et al. [110] noted that overall complication rates from 2006 to 2015 decreased from 13.9% over the first 5 years to 5.5% over the last 5 years, though this change was not statistically significant due to the low number of patients and relatively low complication rates over both periods. However, they assert nonetheless that improved surgical techniques have decreased complication rate [110]. These include 1) elevated placement of the bone flap in combination with expansive duraplasty using a dural substitute, 2) meticulous homeostasis of the dura following excision, and 3) suturing of the dural cuff to the craniotomy margin [110]. Generally, most recent studies note complications rates of ~5 to 11% [105, 107, 109, 110] with the exception of one that notes a complication rate of 23% [108]. However, all studies note that none of the complications led to mortalities, whereas only one notes a permanent complication in a single patient. Together, these results suggest that ECoG implantation has become significantly safer over the last couple decades. The risk of implantation remains not insignificant, and can vary across institutions, but these complications are usually temporary and resolvable with medical care. Of note is that though these studies were performed across multiple institutions with hundreds of patients each, only one mortality was observed, suggesting thorough and regular screening

before and after surgery may significantly mitigate the risks of a chronic implant.

Many of the aforementioned studies have also examined the complication profile for different electrode configurations. Hamer et al. [106] found that complications are associated with a greater number of grids/electrodes (especially > 60 electrodes), longer duration of monitoring (especially > 10 days), older age of the patient, left-sided grid insertion and burr holes in addition to craniotomy. On the other hand, Nagahama et al. [110] found that hemorrhagic and edema/compression complications correlated significantly with the total number of electrode contacts, but not with age, history of prior cranial surgery, laterality, monitoring duration, or number of each electrode type. Although multiple studies agree that the presence of subdural grids [105, 107] or the presence of grids and/or strips generally [108] significantly increases the chance of surgical complication, Vale et al. [105] found no link between greater number of electrodes and increased complication rate. There remains an unsolved discrepancy between the findings of Vale et al. [105] and those of Hamer et al. [106] and Nagahama et al. [110]. However, it is likely that the extent of the implant is more important than the exact number of electrodes, particularly when considering higher-density ECoG grids. This suggests that chronic neural speech decoding studies must strike a balance between minimizing risk and obtaining sufficient coverage for decoding.

Challenges in Translation

ECoG Implantation

The surgical implantation of ECoG electrodes involves breaching the skull in different ways, depending on the extent of the implant. Linear electrode strips can be inserted via burr holes, whereas rectangular grids require craniotomies [111]. Craniotomies carry a higher risk of infection [112, 113]. ECoG grids, though less invasive in some ways than penetrating electrodes, are relatively large and span a larger cortical area. Because of their large surface area, they exhibit a heightened risk of mechanical damage to cortex, due particularly to their mechanical mismatch with brain tissue, relative inflexibility, and inability to mold to contours of the brain [114]. Multiple improvements, such as smaller-profile grids or grids made with more biocompatible materials could minimize the brain's foreign body reaction. Similarly, a combination of novel surgical techniques and electrode designs could mitigate some of the risks of current surgical techniques, such as insertion of expandable or flexible ECoG arrays through smaller burr holes. The relative inflexibility of ECoG grids also limits their ability to image neuronal activity in the sulci of the brain, which may contain valuable information for decoding [114, 115]. Research in this area is progressing steadily as

researchers work to develop such flexible ECoG grids [116–118].

Another issue is the density of electrodes used to sample an area. Standard spacing for ECoG grids is around 10 mm and for high-density grids around 4 mm [119]. These spacings are still too sparse to maximally resolve activity from underlying neural populations, as has been confirmed in a neural speech decoding study by Muller et al. [120], in which information could be gleaned from spacings lower than 4 mm. μ -EGoG arrays with smaller 1-mm spacings have also been used, though primarily for research purposes [121]. However, this increase in resolution comes at the expense of coverage. An optimal ECoG grid should be dense but extensive enough to sample activity from large swaths of cortex. However, an optimum balance in minimizing spacing must be made, as activity from individual electrodes becomes significantly noisier, more difficult to resolve or informationally redundant for BCI applications [122]. A combination of studies suggests that the minimum size for electrodes should be \sim 1 mm and that spacing for subdural and epidural electrodes should be \sim 1.25 and \sim 1.4 mm, respectively [20, 122, 123].

Current ECoG systems require patients to maintain a wired connection with a head stage, amplifiers and a computer when using their BCI. A percutaneous connector, such as the Blackrock NeuroPort, can provide a single connection for 128 or more electrodes [124]. Such an implant also carries a small risk of infection, which is heightened when a patient remains implanted over an extended period of time, and so requires frequent cleaning and monitoring. Nevertheless, the NeuroPort has been demonstrated to be safe in MEA–BCI clinical trials, including BrainGate, Revolutionizing Prosthetics, and NeuroLife [125–127]. BrainGate has implanted several patients at different institutions [125], whereas NeuroLife has worked with a single patient implanted for 5 years now, as of the time of writing [127]. Nonetheless, a wireless ECoG system would drastically reduce the risk of infection by allowing the surgical wound to fully close and heal. This much-awaited advance is on the near horizon; multiple researchers are working towards developing such wireless ECoG systems [118, 128, 129]. However, the drawbacks of wireless systems must also be considered, such as the challenges of transferring large quantities of data wirelessly and doing so at sufficiently high speeds, as well as the issue of repairing such systems when they fail. In order to diagnose any issues with the hardware, it may be necessary to reopen the skull, reintroducing the patient to surgical risk.

Neural Decoding Performance

A speech–BCI system must be able to decode user intention *accurately* enough to allow for fluid communication. Looking to the automatic speech recognition (ASR) literature, early commercial ASR systems in the 1970s obtained word error rates

varying anywhere from 3 to 45%, though these were limited to relatively small data sets of clean, noise-free recordings of speech from single speakers [130]. The performance of ASR systems has increased drastically since then, such as Microsoft’s ASR system, which recently reached a word error rate of 5.8 and 11.0% on the Switchboard and CallHome datasets (each of which contains dozens of hours of speech) matching human-level performance at 5.9 and 11.3% [131]. It is likely that speech BCI will follow a similar trajectory of development. Progress in this area will likely come from development of machine learning algorithms better suited for neural decoding and increased understanding how the brain encodes speech. However, the greatest challenge lies in the relatively small datasets currently used in most neural speech decoding studies, as well as the lack of a standardized dataset in human speech–ECoG with which to objectively benchmark these algorithms. Large datasets have been crucial to the success of deep learning models in ASR [132]. Thus, it is likely that chronic and otherwise large-scale studies will be necessary to collect enough data to allow the application of cutting-edge techniques like deep learning or the creation of a standardized corpus for speech–ECoG.

A speech–BCI system would also require reasonably *low latency* in displaying or synthesizing the decoded speech. Real-time neural decoding is somewhat of a misnomer, as it implies that neural activity is being recorded and decoded simultaneously. Computations take a finite time to complete, and the more complex and numerous the computations, the higher the computational overhead. Therefore, the computational complexity of the decoding and signal-processing scheme must be minimized to reduce computational overhead and allow as close to “real-time” neural decoding as possible. A variety of sources inform what a reasonable target for this may be. In voice communication, 100 to 300 ms of propagation delay causes slight hesitation in a partner’s response and beyond 300 ms causes users to begin to back off to avoid interruption [133]. In delayed auditory feedback, studies note disruptive effects on speech production in subjects with delays starting around 50 ms and maxing out around 200 ms [134, 135], with delays of 150 ms or so even inducing stress [136]; however, disruptive effects occur even with delays as low as 25 ms when the subject is instructed to speak quickly [134, 135]. Together, these results suggest a \sim 50-ms latency target for an ideal speech BCI system, though latencies of up to \sim 100 ms or so may still be reasonable for facilitating communication.

An alternative way to examine latency in speech–BCI systems is to determine the *maximum information transfer rate* needed for speech BCI to serve as an input interface, as latency would necessarily decrease the maximum number of words or phonemes that could be transmitted. One study cites an average (averaged from both studies) and maximum human typing speed of 57.4 and 104 words per minute, respectively [137]. Using these numbers, along with estimates that English words contain about 4.79 letters on average [138] and that

English phonemes correspond to ~ 1.88 letters (not accounting for frequency) [139], this results in latencies per phoneme of about 410 or 226 ms. Similarly, another study cites an average and maximum words per minute during human conversation of 196 and 291 words per minute, respectively [140]. With the previous assumptions, these correspond to latencies per phoneme of about 120 or 81 ms. Together, these suggest an ideal latency target of ~ 81 ms for speech BCI, though ~ 120 ms latency would be sufficient for average conversational speech and latency in the 100s of milliseconds would be sufficient for communicating by typing, which aligns reasonably with the previous findings.

Finally, though deficiencies in this area may be better tolerated by locked-in patients, an ideal speech–BCI system would likely need to be *generalizable* between users and require as little individualized training of the system for each new user beforehand or additional calibration following. BCI studies to date typically require extensive training over a period of several days or weeks, typically with a 20- to 30-min recalibration being necessary prior to each session [141]. However, recent studies suggest that initial training time can be reduced to less than a minute and that subsequent calibration time can be eliminated entirely [141–143]. Replicating such low initial training and subsequent calibration times in speech BCI would be highly desirable. Only a couple decades ago, it took a state-of-the-art ASR speech-to-text dictation system, “Dragon Speaking Naturally,” 20 min of speech read by the user plus 10 min of processing time, followed by the user feeding in several of his or her previously typed documents to tailor the system to the user’s vocabulary and 2.5 h of usage with manual corrections, before the system could achieve 70 to 80% accuracy [144]. Today, smartphones come equipped with personal voice assistants like Siri or Google Voice, which require no training on the part of the user, obtain results from their servers in a fraction of a second and work with almost unlimited vocabularies [145] with word error rates as low as 4.9% [146]. Until speech–BCI systems approach the nonexistent training times of current personal voice assistants, it is unlikely that they will be widely adopted. However, this is a challenging area of research, complicated by the fact that neural representations of speech can appear so differently within and between individuals, even more so than in traditional speech recognition. Nonetheless, intensified efforts in this area and further advances in more robust or adaptable machine learning and transfer learning may pave the way towards overcoming this major challenge.

Conclusions

This review attempts a comprehensive overview of how speech is represented in the brain and how it might be decoded in a BCI. The key approaches to speech BCI arise in auditory,

articulatory, and semantic neural speech decoding, each enabling a fundamentally different but complementary form of speech BCI. However, it may be that a multimodal approach combining these different types of speech information will be necessary to create an optimally performing speech BCI. We also outline herein the reasons why ECoG may be an ideal neural recording modality for speech BCI. A growing literature is demonstrating not only that real-time neural speech decoding with ECoG is possible but also that ECoG can facilitate a relatively safe and stable, long-term BCI. Though multiple barriers for widespread translation of ECoG-based speech BCIs exist, such devices could potentially help patients with locked-in syndrome and other severe communication disorders in the short-term, allowing them to participate more effectively and fully in their lives. With further advances in the safety and efficacy of both invasive and noninvasive BCIs, this technology may also reach a broader population.

Acknowledgments: The authors of this paper have been supported by the National Institutes of Health (R01 NS091139).

Required Author Forms Disclosure forms provided by the authors are available with the online version of this article.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, et al. Classification of primary progressive aphasia and its variants. *Neurology* 2011;76(11):1006–14.
- Smith E, Delargy M. Locked-in syndrome. *BMJ* 2005;330(7488):406–9.
- Indefrey P. The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Front Psychol* [Internet]. 2011;2. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3191502/>
- Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. *Hum Brain Mapp* 1994;1(2):153–71.
- Irani F, Platek SM, Bunce S, Ruocco AC, Chute D. Functional near infrared spectroscopy (fNIRS): an emerging neuroimaging technology with important applications for the study of brain disorders. *Clin Neuropsychol* 2007;21(1):9–37.
- Dassios G, Fokas AS, Kariotou F. On the non-uniqueness of the inverse MEG problem. *Inverse Probl* 2005;21(2):L1.
- Baillet S, Mosher JC, Leahy RM. Electromagnetic brain mapping. *IEEE Signal Process Mag* 2001;18(6):14–30.
- Gotman J. High frequency oscillations: the new EEG frontier? *Epilepsia* 2010;51(Suppl 1):63–5.
- Michel CM, Murray MM. Towards the utilization of EEG as a brain imaging tool. *NeuroImage* 2012;61(2):371–85.
- Zelmann R, Lina JM, Schulze-Bonhage A, Gotman J, Jacobs J. Scalp EEG is not a blur: it can see high frequency oscillations although their generators are small. *Brain Topogr* 2014;27(5):683–704.

11. Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, et al. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *J Neural Eng* 2018;15(3):031005.
12. Maynard EM, Nordhausen CT, Normann RA. The Utah Intracortical Electrode Array: a recording structure for potential brain–computer interfaces. *Electroencephalogr Clin Neurophysiol* 1997;102(3):228–39.
13. Biran R, Martin DC, Tresco PA. Neuronal cell loss accompanies the brain tissue response to chronically implanted silicon microelectrode arrays. *Exp Neurol* 2005;195(1):115–26.
14. Prodanov D, Delbecke J. Mechanical and biological interactions of implants with the brain and their impact on implant design. *Front Neurosci* [Internet]. 2016;10 [cited 2018 Jun 5]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4746296/>
15. Hochberg LR, Serruya MD, Friehs GM, Mukand JA, Saleh M, Caplan AH, et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 2006;442(7099):164–71.
16. Bouton CE, Shaikhoui A, Annetta NV, Bockbrader MA, Friedenber DA, Nielson DM, et al. Restoring cortical control of functional movement in a human with quadriplegia. *Nature* 2016;533(7602):247–50.
17. Collinger JL, Wodlinger B, Downey JE, Wang W, Tyler-Kabara EC, Weber DJ, et al. High-performance neuroprosthetic control by an individual with tetraplegia. *Lancet* 2013;381(9866):557–64.
18. Crone NE, Sinai A, Korzeniewska A. High-frequency gamma oscillations and human brain mapping with electrocorticography. In: Neuper C, Klimesch W, editors. *Progress in Brain Research* [Internet]. Elsevier; 2006. p. 275–95. (Event-Related Dynamics of Brain Oscillations; vol. 159). [cited 2018 Aug 30] Available from: <http://www.sciencedirect.com/science/article/pii/S0079612306590193>
19. Milsap G, Collard M, Coogan C, Crone NE. BCI2000Web and WebFM: browser-based tools for brain computer interfaces and functional brain mapping. *Front Neurosci* 2019
20. Slutzky MW, Jordan LR, Krieg T, Chen M, Mogul DJ, Miller LE. Optimal spacing of surface electrode arrays for brain machine interface applications. *J Neural Eng* 2010;7(2):26004.
21. Rupp K, Roos M, Milsap G, Caceres C, Ratto C, Chevillet M, et al. Semantic attributes are encoded in human electrocorticographic signals during visual object recognition. *NeuroImage* 2017;148:318–29.
22. Huth AG, de Heer WA, Griffiths TL, Theunissen FE, Gallant JL. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 2016;532(7600):453–8.
23. Mines MA, Hanson BF, Shoup JE. Frequency of occurrence of phonemes in conversational English. *Lang Speech* 1978;21(3):221–41.
24. Hickok G, Poeppel D. Chapter 25—neural basis of speech perception. In: Hickok G, Small SL, editors. *Neurobiology of Language* [Internet]. San Diego: Academic Press; 2016. p. 299–310. [cited 2018 Aug 30] Available from: <http://www.sciencedirect.com/science/article/pii/B9780124077942000250>
25. Titze IR. *Principles of Voice Production*. Prentice Hall; 1994. 390 p.
26. Gasser M. 3.3 Vowels. In: *How Language Works: The Cognitive Science of Linguistics* [Internet]. 2012 [cited 2017 Dec 24]. Available from: <http://www.indiana.edu/~hlw/PhonUnits/vowels.html>
27. Hickok G, Houde J, Rong F. Sensorimotor Integration in speech processing: computational basis and neural organization. *Neuron* 2011;69(3):407–22.
28. Conant D, Bouchard KE, Chang EF. Speech map in the human ventral sensory-motor cortex. *Curr Opin Neurobiol* 2014;24:63–7.
29. Rubin P, Baer T, Mermelstein P. An articulatory synthesizer for perceptual research. *J Acoust Soc Am* 1981;70(2):321–8.
30. Uria B, Murray I, Renals S, Richmond K. Deep architectures for articulatory inversion [Internet]. INTERSPEECH-2012; 2012. Available from: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.248.6643>
31. Hotz-Boendemaker S, Funk M, Summers P, Brugger P, Hepp-Reymond M-C, Curt A, et al. Preservation of motor programs in paraplegics as demonstrated by attempted and imagined foot movements. *NeuroImage* 2008 Jan 1;39(1):383–94.
32. Hermes D, Vansteensel MJ, Albers AM, Bleichner MG, Benedictus MR, Orellana CM, et al. Functional MRI-based identification of brain areas involved in motor imagery for implantable brain–computer interfaces. *J Neural Eng* 2011;8(2):025007.
33. Liu H, Agam Y, Madsen JR, Kreiman G. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 2009;62(2):281–90.
34. Wang W, Degenhart AD, Sudre GP, Pomerleau DA, Tyler-Kabara EC. Decoding semantic information from human electrocorticographic (ECoG) signals. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011. p. 6294–8.
35. Manning JR, Sperling MR, Sharan A, Rosenberg EA, Kahana MJ. Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *J Neurosci* 2012;32(26):8871–8.
36. Chen Y, Shimotake A, Matsumoto R, Kunieda T, Kikuchi T, Miyamoto S, et al. The ‘when’ and ‘where’ of semantic coding in the anterior temporal lobe: temporal representational similarity analysis of electrocorticogram data. *Cortex* 2016;79:1–13.
37. Moerel M, De Martino F, Formisano E. An anatomical and functional topography of human auditory cortical areas. *Front Neurosci* [Internet]. 2014;8. [cited 2017 Nov 13]. Available from: <https://www.frontiersin.org/articles/10.3389/fnins.2014.00225/full#h4>
38. Humphries C, Liebenthal E, Binder JR. Tonotopic organization of human auditory cortex. *NeuroImage* 2010;50(3):1202–11.
39. Langers DRM, van Dijk P. Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. *Cereb Cortex* 2012;22(9):2024–38.
40. Su L, Zulfqar I, Jamshed F, Fonteneau E, Marslen-Wilson W. Mapping tonotopic organization in human temporal cortex: representational similarity analysis in EMEG source space. *Front Neurosci* [Internet] 2014;e8. [cited 2017 Dec 25] Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4228977/>
41. Norman-Haignere S, Kanwisher N, McDermott JH. Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J Neurosci* 2013;33(50):19451–69.
42. Pantev C, Hoke M, Lutkenhoner B, Lehnertz K. Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science* 1989;246(4929):486–8.
43. Mesgarani N, Cheung C, Johnson K, Chang EF. Phonetic feature encoding in human superior temporal gyrus. *Science* 2014;343(6174):1006–10.
44. Hullett PW, Hamilton LS, Mesgarani N, Schreiner CE, Chang EF. Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. *J Neurosci* 2016;36(6):2014–26.
45. Hamilton LS, Edwards E, Chang EF. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr Biol* 2018;28(12):1860–1871.e4.
46. Yao B, Belin P, Scheepers C. Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *J Cogn Neurosci* 2011;23(10):3146–52.
47. Perrone-Bertolotti M, Kujala J, Vidal JR, Hamame CM, Ossandon T, Bertrand O, et al. How silent is silent reading? Intracerebral

- evidence for top-down activation of temporal voice areas during reading. *J Neurosci* 2012;32(49):17554–62.
48. Perrone-Bertolotti M, Pichat C, Le Bas JF, Baciú A, Baciú M. Functional MRI evidence for modulation of cerebral activity by grapheme-to-phoneme conversion in French, and by the variable of gender. *J Neurolinguistics* 2011;24(4):507–20.
 49. Chan AM, Dykstra AR, Jayaram V, Leonard MK, Travis KE, Gygi B, et al. Speech-specific tuning of neurons in human superior temporal gyrus. *Cereb Cortex N Y NY* 2014;24(10):2679–93.
 50. Magrassi L, Aromataris G, Cabrini A, Annovazzi-Lodi V, Moro A. Sound representation in higher language areas during language generation. *Proc Natl Acad Sci* 2015;112(6):1868–73.
 51. Brumberg JS, Krusienski DJ, Chakrabarti S, Gunduz A, Brunner P, Ritaccio AL, et al. Spatio-temporal progression of cortical activity related to continuous overt and covert speech production in a reading task. *PLoS One* 2016 Nov 22;11(11):e0166872.
 52. Martin S, Brunner P, Holdgraf C, Heinze H-J, Crone NE, Rieger J, et al. Decoding spectrotemporal features of overt and covert speech from the human cortex. *Front Neuroengineering* [Internet]. 2014;7. [cited 2017 Dec 19] Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4034498/>
 53. 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.
 54. Leuthardt EC, Pei X-M, Breshears J, Gaona C, Sharma M, Freudenberg Z, et al. Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task. *Front Hum Neurosci* [Internet]. 2012;6. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3342676/>
 55. Tian X, Poeppel D. Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front Psychol* [Internet]. 2010;1. [cited 2018 Jul 11] Available from: <https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00166/full>
 56. 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.
 57. Ikeda S, Shibata T, Nakano N, Okada R, Tsuyuguchi N, Ikeda K, et al. Neural decoding of single vowels during covert articulation using electrocorticography. *Front Hum Neurosci* [Internet]. 2014;8. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3945950/>
 58. Penfield W, Boldrey E. Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*. 1937;60(4):389–443.
 59. Schieber MH. Constraints on somatotopic organization in the primary motor cortex. *J Neurophysiol* 2001;86(5):2125–43.
 60. Donoghue JP, Nurmikko A, Black M, Hochberg LR. Assistive technology and robotic control using motor cortex ensemble-based neural interface systems in humans with tetraplegia. *J Physiol* 2007;579(3):603–11.
 61. Hotson G, McMullen DP, Fifer MS, Johannes MS, Katyal KD, Para MP, et al. Individual finger control of the modular prosthetic limb using high-density electrocorticography in a human subject. *J Neural Eng* 2016;13(2):026017.
 62. Sharma G, Annetta N, Friedenber DA, Bockbrader M. Advances in BCI: a neural bypass technology to reconnect the brain to the body. In: *Brain-Computer Interface Research* [Internet]. Springer, Cham; 2017 [cited 2017 Nov 13]. p. 9–20. (SpringerBriefs in Electrical and Computer Engineering). Available from: https://link.springer.com/chapter/10.1007/978-3-319-64373-1_2
 63. Bouchard KE, Mesgarani N, Johnson K, Chang EF. Functional organization of human sensorimotor cortex for speech articulation. *Nature* 2013;495(7441):327–32.
 64. Toyoda G, Brown EC, Matsuzaki N, Kojima K, Nishida M, Asano E. Electro-corticographic correlates of overt articulation of 44 English phonemes: intracranial recording in children with focal epilepsy. *Clin Neurophysiol* 2014;125(6):1129–37.
 65. Lotte F, Brumberg JS, Brunner P, Gunduz A, Ritaccio AL, Guan C, et al. Electro-corticographic representations of segmental features in continuous speech. *Front Hum Neurosci* [Internet]. 2015;9. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4338752/>
 66. Cheung C, Hamilton LS, Johnson K, Chang EF. The auditory representation of speech sounds in human motor cortex. *elife* 2016;5:e12577.
 67. Arsenault JS, Buchsbaum BR. No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception. *Psychon Bull Rev* 2016;23(4):1231–40.
 68. Glanz Iljina O, Derix J, Kaur R, Schulze-Bonhage A, Auer P, Aertsen A, et al. Real-life speech production and perception have a shared premotor-cortical substrate. *Sci Rep* 2018;8(1):8898.
 69. Tankus A, Fried I, Shoham S. Structured neuronal encoding and decoding of human speech features. *Nat Commun* 2012;3:ncmms1995.
 70. Mugler EM, Patton JL, Flint RD, Wright ZA, Schuele SU, Rosenow J, et al. Direct classification of all American English phonemes using signals from functional speech motor cortex. *J Neural Eng* 2014;11(3):035015.
 71. Mugler EM, Goldrick M, Rosenow JM, Tate MC, Slutzky MW. Decoding of articulatory gestures during word production using speech motor and premotor cortical activity. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015. p. 5339–42.
 72. Bouchard KE, Chang EF. Neural decoding of spoken vowels from human sensory-motor cortex with high-density electrocorticography. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014. p. 6782–5.
 73. 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.
 74. Mugler EM, Goldrick M, Slutzky MW. Cortical encoding of phonemic context during word production. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014. p. 6790–3.
 75. 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. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2012. p. 3292–5.
 76. Crone NE, Sinai A, Korzeniewska A. High-frequency gamma oscillations and human brain mapping with electrocorticography. In: Neuper C, Klimesch W, editors. *Progress in Brain Research* [Internet]. Elsevier; 2006 [cited 2017 Dec 3]. p. 275–95. (Event-Related Dynamics of Brain Oscillations; vol. 159). Available from: <http://www.sciencedirect.com/science/article/pii/S0079612306590193>
 77. Buzsáki G, Wang X-J. Mechanisms of gamma oscillations. *Annu Rev Neurosci* 2012;35:203–25.
 78. Yazdan-Shahmorad A, Kipke DR, Lehmkuhle MJ. High gamma power in ECoG reflects cortical electrical stimulation effects on unit activity in layers V/VI. *J Neural Eng* 2013;10(6):066002.
 79. Martin S, Iturrate I, del Millán JR, Knight RT, Pasley BN. Decoding inner speech using electrocorticography: progress and challenges toward a speech prosthesis. *Front Neurosci* [Internet]. 2018 ;12 [cited 2018 Jul 18]. Available from: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00422/full>

80. Vansteensel MJ, Pels EGM, Bleichner MG, Branco MP, Denison T, Freudenburg ZV, et al. Fully implanted brain–computer interface in a locked-in patient with ALS. *N Engl J Med* 2016;375(21):2060–6.
81. Guenther FH, Brumberg JS, Wright EJ, Nieto-Castanon A, Tourville JA, Panko M, et al. A wireless brain-machine interface for real-time speech synthesis. *PLoS One* 2009;4(12):e8218.
82. Herff C, Heger D, de Pestera A, Telaar D, Brunner P, Schalk G, et al. Brain-to-text: decoding spoken phrases from phone representations in the brain. *Front Neurosci* [Internet]. 2015;9. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4464168/>
83. Herff C, de Pestera A, Heger D, Brunner P, Schalk G, Schultz T, et al. Towards continuous speech recognition for BCI. 2017;21–9.
84. Moses DA, Mesgarani N, Leonard MK, Chang EF. Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. *J Neural Eng* 2016;13(5):056004.
85. Moses DA, Leonard MK, Chang EF. Real-time classification of auditory sentences using evoked cortical activity in humans. *J Neural Eng* 2018;15(3):036005.
86. Bruns A. Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? *J Neurosci Methods* 2004;137(2):321–32.
87. Bocquelet F, Hueber T, Girin L, Savariaux C, Yvert B. Real-time control of an articulatory-based speech synthesizer for brain computer interfaces. *PLoS Comput Biol* 2016;12(11):e1005119.
88. Bouchard KE, Conant DF, Anumanchipalli GK, Dichter B, Chaisanguanthum KS, Johnson K, et al. High-resolution, non-invasive imaging of upper vocal tract articulators compatible with human brain recordings. *PLoS One* 2016;11(3):e0151327.
89. Conant DF, Bouchard KE, Leonard MK, Chang EF. Human sensorimotor cortex control of directly measured vocal tract movements during vowel production. *J Neurosci* 2018;38(12):2955–66.
90. Herff C, Johnson G, Diener L, Shih J, Krusienski D, Schultz T. Towards direct speech synthesis from ECoG: a pilot study. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf* 2016;2016:1540–3.
91. Akbari H, Khalighinejad B, Herrero J, Mehta A, Mesgarani N. Reconstructing intelligible speech from the human auditory cortex. *bioRxiv*. 2018 19;350124.
92. Milsap G, Collard M, Coogan C, Rabbani Q, Wang Y, Nathan E. C. Keyword spotting using human electrocorticographic recordings. *Front Neurosci*. 2019
93. Yeager JD, Phillips DJ, Rector DM, Bahr DF. Characterization of flexible ECoG electrode arrays for chronic recording in awake rats. *J Neurosci Methods* 2008;173(2):279–85.
94. Chao ZC, Nagasaka Y, Fujii N. Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkey. *Front Neuroengineering* [Internet]. 2010;3. [cited 2017 Jul 8] Available from: <http://journal.frontiersin.org/article/10.3389/fneng.2010.00003/full>
95. Rouse AG, Williams JJ, Wheeler JJ, Moran DW. Cortical adaptation to a chronic micro-electrocorticographic brain computer interface. *J Neurosci* 2013;33(4):1326–30.
96. Degenhart AD, Eles J, Dum R, Mischel JL, Smalianchuk I, Eandler B, et al. Histological evaluation of a chronically-implanted electrocorticographic electrode grid in a non-human primate. *J Neural Eng* 2016;13(4):046019.
97. Duckrow RB, Teheng TK. Daily variation in an intracranial EEG feature in humans detected by a responsive neurostimulator system. *Epilepsia* 2007;48(8):1614–20.
98. Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. *Lancet Neurol* 2013;12(6):563–71.
99. Spencer DC, Sun FT, Brown SN, Jobst BC, Fountain NB, Wong VSS, et al. Circadian and ultradian patterns of epileptiform discharges differ by seizure-onset location during long-term ambulatory intracranial monitoring. *Epilepsia* 2016;57(9):1495–502.
100. Smart O, Rolston JD, Epstein CM, Gross RE. Hippocampal seizure-onset laterality can change over long timescales: a same-patient observation over 500 days. *Epilepsy Behav Case Rep* 2013;1:56–61.
101. Rao VR, Leonard MK, Kleen JK, Lucas BA, Mirro EA, Chang EF. Chronic ambulatory electrocorticography from human speech cortex. *NeuroImage* 2017;153:273–82.
102. Bundy DT, Zellmer E, Gaona CM, Sharma M, Szrama N, Hacker C, et al. Characterization of the effects of the human dura on macro- and micro-electrocorticographic recordings. *J Neural Eng* 2014;11(1):016006.
103. Flint RD, Rosenow JM, Tate MC, Slutzky MW. Continuous decoding of human grasp kinematics using epidural and subdural signals. *J Neural Eng* 2016;14(1):016005.
104. Leuthardt EC, Gaona C, Sharma M, Szrama N, Roland J, Freudenberg Z, et al. Using the electrocorticographic speech network to control a brain-computer interface in humans. *J Neural Eng* 2011;8(3):036004.
105. Vale FL, Pollock G, Dionisio J, Benbadis SR, Tatum WO. Outcome and complications of chronically implanted subdural electrodes for the treatment of medically resistant epilepsy. *Clin Neurol Neurosurg Assen* 2013;115(7):985–90.
106. Hamer HM, Morris HH, Mascha EJ, Karafa MT, Bingaman WE, Bej MD, et al. Complications of invasive video-EEG monitoring with subdural grid electrodes. *Neurology* 2002;58(1):97–103.
107. Hedegård E, Bjellvi J, Edelvik A, Rydenhag B, Flink R, Malmgren K. Complications to invasive epilepsy surgery workup with subdural and depth electrodes: a prospective population-based observational study. *J Neurol Neurosurg Psychiatry* 2014;85(7):716–20.
108. Wellmer J, von der Groeben F, Klarman U, Weber C, Elger CE, Urbach H, et al. Risks and benefits of invasive epilepsy surgery workup with implanted subdural and depth electrodes. *Epilepsia* 2012;53(8):1322–32.
109. Mathon B, Clemenceau S, Hasboun D, Habert M-O, Belaid H, Nguyen-Michel V-H, et al. Safety profile of intracranial electrode implantation for video-EEG recordings in drug-resistant focal epilepsy. *J Neurol* 2015;262(12):2699–712.
110. Nagahama Y, Schmitt AJ, Nakagawa D, Vesole AS, Kamm J, Kovach CK, et al. Intracranial EEG for seizure focus localization: evolving techniques, outcomes, complications, and utility of combining surface and depth electrodes. *J Neurosurg* 2018;1(aop):1–13.
111. Voorhies JM, Cohen-Gadol A. Techniques for placement of grid and strip electrodes for intracranial epilepsy surgery monitoring: pearls and pitfalls. *Surg Neurol Int* [Internet] 2013;4. [cited 2017 Dec 3] Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3740610/>
112. Kourbeti IS, Vakis AF, Ziakas P, Karabetos D, Potolidis E, Christou S, et al. Infections in patients undergoing craniotomy: risk factors associated with post-craniotomy meningitis. *J Neurosurg* 2014;122(5):1113–9.
113. Patel KS, Goldenberg B, Schwartz TH. Betadine irrigation and post-craniotomy wound infection. *Clin Neurol Neurosurg* 2014;118:49–52.
114. Weltman A, Yoo J, Meng E. Flexible, penetrating brain probes enabled by advances in polymer microfabrication. *Micromachines* 2016;7(10):180.
115. Matsuo T, Kawasaki K, Osada T, Sawahata H, Suzuki T, Shibata M, et al. Intracranial electrocorticography in macaque monkeys with minimally invasive neurosurgical protocols. *Front Syst Neurosci* [Internet]. 2011;5. [cited 2017 Dec 5] Available from: <https://www.frontiersin.org/articles/10.3389/fnsys.2011.00034/full>
116. Castagnola E, Maiolo L, Maggolini E, Minotti A, Marrani M, Maita F, et al. PEDOT-CNT-coated low-impedance, ultra-flexible,

- and brain-conformable micro-ECoG arrays. *IEEE Trans Neural Syst Rehabil Eng* 2015;23(3):342–50.
117. Kaiju T, Doi K, Yokota M, Watanabe K, Inoue M, Ando H, et al. High spatiotemporal resolution ECoG recording of somatosensory evoked potentials with flexible micro-electrode arrays. *Front Neural Circuits* [Internet]. 2017;11. [cited 2017 Dec 3] Available from: <https://www.frontiersin.org/articles/10.3389/fncir.2017.00020/full>
 118. Xie K, Zhang S, Dong S, Li S, Yu C, Xu K, et al. Portable wireless electrocorticography system with a flexible microelectrodes array for epilepsy treatment. *Sci Rep* 2017;7(1):7808.
 119. Wang PT, King CE, McCrimmon CM, Lin JJ, Sazgar M, Hsu FPK, et al. Comparison of decoding resolution of standard and high-density electrocorticogram electrodes. *J Neural Eng* 2016;13(2):026016.
 120. Muller L, Hamilton LS, Edwards E, Bouchard KE, Chang EF. Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography. *J Neural Eng* 2016;13(5):056013.
 121. Thongpang S, Richner TJ, Brodnick SK, Schendel A, Kim J, Wilson JA, et al. A micro-electrocorticography platform and deployment strategies for chronic BCI applications. *Clin EEG Neurosci* 2011;42(4):259–65.
 122. Wodlinger B, Degenhart AD, Collinger JL, Tyler-Kabara EC, Wang W. The impact of electrode characteristics on electrocorticography (ECoG). In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011. p. 3083–6.
 123. Freeman WJ, Rogers LJ, Holmes MD, Silbergeld DL. Spatial spectral analysis of human electrocorticograms including the alpha and gamma bands. *J Neurosci Methods* 2000;95(2):111–21.
 124. Blackrock Microsystems [Internet]. [cited 2018 Jul 20]. Available from: <http://blackrockmicro.com/>
 125. BrainGate [Internet]. BrainGate. [cited 2018 Nov 6]. Available from: <https://www.braingate.org/>
 126. Revolutionizing prosthetics [Internet]. JHUAPL. [cited 2018 Nov 6]. Available from: <https://www.jhuapl.edu/prosthetics/>
 127. NeuroLife neural bypass technology [Internet]. Battelle. [cited 2018 Nov 6]. Available from: <https://www.battelle.org/government-offerings/health/medical-devices/neurotechnology/neurolife-neural-bypass-technology>
 128. Matsushita K, Hirata M, Suzuki T, Ando H, Ota Y, Sato F, et al. Development of an implantable wireless ECoG 128ch recording device for clinical brain machine interface. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2013. p. 1867–70.
 129. Yoshimoto S, Araki T, Uemura T, Nezu T, Sekitani T, Suzuki T, et al. Implantable wireless 64-channel system with flexible ECoG electrode and optogenetics probe. In: 2016 IEEE Biomedical Circuits and Systems Conference (BioCAS). 2016. p. 476–9.
 130. Reddy DR. Speech recognition by machine: a review. *Proc IEEE* 1976;64(4):501–31.
 131. Xiong W, Droppo J, Huang X, Seide F, Seltzer M, Stolcke A, et al. Achieving human parity in conversational speech recognition. *ArXiv161005256 Cs* [Internet]. 2016 [cited 2017 Dec 4]; Available from: <http://arxiv.org/abs/1610.05256>
 132. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
 133. Na S, Yoo S. Allowable propagation delay for VoIP calls of acceptable quality. In: *Advanced Internet Services and Applications* [Internet]. Springer, Berlin, Heidelberg; 2002 [cited 2017 Dec 4]. p. 47–55. (Lecture Notes in Computer Science). Available from: https://link.springer.com/chapter/10.1007/3-540-45639-2_6
 134. Stuart A, Kalinowski J, Rastatter MP, Lynch K. Effect of delayed auditory feedback on normal speakers at two speech rates. *J Acoust Soc Am* 2002;111(5):2237–41.
 135. Stuart A, Kalinowski J. Effect of delayed auditory feedback, speech rate, and sex on speech production. *Percept Mot Skills* 2015;120(3):747–65.
 136. Badian M, Appel E, Palm D, Rupp W, Sittig W, Taeuber K. Standardized mental stress in healthy volunteers induced by delayed auditory feedback (DAF). *Eur J Clin Pharmacol* 1979;16(3):171–6.
 137. Salthouse TA. Effects of age and skill in typing. *J Exp Psychol Gen* 1984;113(3):345–71.
 138. Norvig P. English Letter frequency counts: Mayzner Revisited or ETAOIN SRHLDCU [Internet]. [cited 2018 Jun 15]. Available from: <http://norvig.com/mayzner.html>
 139. The 44 sounds (phonemes) of English [Internet]. DSF Literacy Resources; [cited 2017 Dec 17]. Available from: <https://www.dyslexia-reading-well.com/support-files/the-44-phonemes-of-english.pdf>
 140. Yuan J, Liberman M, Cieri C. Towards an integrated understanding of speaking rate in conversation. In: *Conversation Interspeech*. 2006. p. 541–544.
 141. Krauledat M, Schröder M, Blankertz B, Müller K-R. Reducing calibration time for brain-computer interfaces: a clustering approach. In: *Advances in Neural Information Processing Systems* 19. MIT Press; 2007. p. 753–760.
 142. Brandman DM, Hosman T, Saab J, Burkhart MC, Shanahan BE, Ciancibello JG, et al. Rapid calibration of an intracortical brain–computer interface for people with tetraplegia. *J Neural Eng* 2018;15(2):026007.
 143. Verhoeven T, Hübner D, Tangermann M, Müller KR, Dambre J, Kindermans PJ. Improving zero-training brain-computer interfaces by mixing model estimators. *J Neural Eng* 2017;14(3):036021.
 144. Hedberg SR. Dictating this article to my computer: automatic speech recognition is coming of age. *IEEE Expert* 1997;12(6):9–11.
 145. Huang X, Baker J, Reddy R. A historical perspective of speech recognition. *Commun ACM* 2014;57(1):94–103.
 146. Protalinski E. Google’s speech recognition technology now has a 4.9% word error rate [Internet]. *VentureBeat*. 2017 [cited 2017 Dec 5]. Available from: <https://venturebeat.com/2017/05/17/googles-speech-recognition-technology-now-has-a-4-9-word-error-rate/>