Neurotherapeutics
https://doi.org/10.1007/s13311-022-01190-2

CURRENT PERSPECTIVES

Brain-Computer Interface: Applications to Speech Decoding and Synthesis to Augment Communication

Shiyu Luo[1] · **Qinwan Rabbani**[2] · **Nathan E. Crone**[3]

Accepted: 16 January 2022
© The American Society for Experimental NeuroTherapeutics, Inc. 2022

Abstract
Damage or degeneration of motor pathways necessary for speech and other movements, as in brainstem strokes or amyotrophic lateral sclerosis (ALS), can interfere with efficient communication without affecting brain structures responsible for language or cognition. In the worst-case scenario, this can result in the locked in syndrome (LIS), a condition in which individuals cannot initiate communication and can only express themselves by answering yes/no questions with eye blinks or other rudimentary movements. Existing augmentative and alternative communication (AAC) devices that rely on eye tracking can improve the quality of life for people with this condition, but brain-computer interfaces (BCIs) are also increasingly being investigated as AAC devices, particularly when eye tracking is too slow or unreliable. Moreover, with recent and ongoing advances in machine learning and neural recording technologies, BCIs may offer the only means to go beyond cursor control and text generation on a computer, to allow real-time synthesis of speech, which would arguably offer the most efficient and expressive channel for communication. The potential for BCI speech synthesis has only recently been realized because of seminal studies of the neuroanatomical and neurophysiological underpinnings of speech production using intracranial electrocorticographic (ECoG) recordings in patients undergoing epilepsy surgery. These studies have shown that cortical areas responsible for vocalization and articulation are distributed over a large area of ventral sensorimotor cortex, and that it is possible to decode speech and reconstruct its acoustics from ECoG if these areas are recorded with sufficiently dense and comprehensive electrode arrays. In this article, we review these advances, including the latest neural decoding strategies that range from deep learning models to the direct concatenation of speech units. We also discuss state-of-the-art vocoders that are integral in constructing natural-sounding audio waveforms for speech BCIs. Finally, this review outlines some of the challenges ahead in directly synthesizing speech for patients with LIS.

Keywords Speech synthesis · Brain-computer interface · Locked-in syndrome · Electrocorticography · ECoG

Background

Clinical Needs

A speech brain-computer interface (BCI) is a method of alternative and augmentative communication (AAC) based on measuring and interpreting neural signals generated during attempted or imagined speech [1, 2]. The greatest need for speech BCI occurs in patients with motor and speech impairments due to acute or degenerative lesions of the pyramidal tracts or lower motor neurons without significant impairment of language or cognition. When movement and speech impairments are particularly severe, as in the locked in syndrome, patients may be unable to independently initiate or sustain communication and may be limited to answering yes/no questions with eye blinks, eye movements, or other minor residual movements. Significant advances have been made to assist these individuals through the use of other types of BCIs, including those using P300 [3], motor imagery [4], handwriting [5], and steady-state visually evoked potential [6]. However, these forms of communication cannot replace the speed and flexibility of spoken communication. The average words communicated per

✉ Shiyu Luo
sluo15@jhu.edu

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

2 Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD, USA

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

Published online: 31 January 2022

minute in conversational speech is more than 7 times that of eye tracking and 10 times of handwriting [7, 8]. Finally, speech allows patients to communicate with less effort as it is a more natural and intuitive modality for information exchange.

Invasive vs Non-invasive BCIs for Speech

Although non-invasive methods of measuring neural activity have been used as a BCI, no existing non-invasive recording method delivers adequate spatial and temporal resolution for use as a speech BCI. Imaging techniques such as functional near-infrared spectroscopy (fNIRS) and functional magnetic resonance imaging (fMRI) provide a delayed and indirect measure of neural activity with low temporal resolution, albeit with relatively good spatial resolution. Although magnetoencephalography (MEG) and electroencephalography (EEG) have adequate temporal resolution, they lack sufficient spatial resolution [9]. Moreover, MEG currently requires a magnetically shielded room, limiting its use to laboratory environments. Although EEG can be recorded at the scalp surface with electrode caps, these caps are cumbersome and require continued attention to electrode impedances to maintain adequate signal quality. Despite their limitations on resolution, fMRI, MEG, and EEG can provide expansive spatial coverage, which is advantageous when investigating the dynamics of the widely distributed language networks.

Because of the limitations of current non-invasive recording techniques, most work on speech BCI has been focused on using electrophysiological recordings of cortical neuronal activity with implanted electrodes of varying sizes and configurations [10]. These recordings have focused either on action potentials generated by single neurons or on local field potentials generated by populations of cortical neurons. Most advances in BCI research have arisen from techniques that record action potentials or related multi-unit activity from an ever-increasing number of microelectrodes. Until recently, the gold standard for these recordings used 2D arrays of up to 128 electrodes, each with single recording tips (Fig. 1). However, recent advances have allowed for up to 32 recording contacts along each implanted electrode, allowing even more single units to be recorded within a small volume of cortical tissue. Robotic operative techniques are also being developed to insert electrodes with less trauma to cortical tissue [11]. These techniques are designed to maximize the number of single units recorded per square millimeter of tissue. However, conventional wisdom that has the native cortical representations for vocalization and articulation during speech is widely distributed over most of the ventral portion of sensorimotor cortex in the pre- and post-central gyrus, and thus, any attempt to leverage these representations in a speech BCI will require recordings that can sample from a large surface area. Despite this, recent studies have shown the possibility of decoding speech from microelectrode Utah arrays implanted in dorsal motor areas [12, 13]. Stereo-electroencephalographic (sEEG) depth arrays have also been suggested as a promising recording modality for speech BCI (see detailed review in [14]). sEEG electrodes are thin depth electrodes surgically implanted through small holes in the skull, which makes

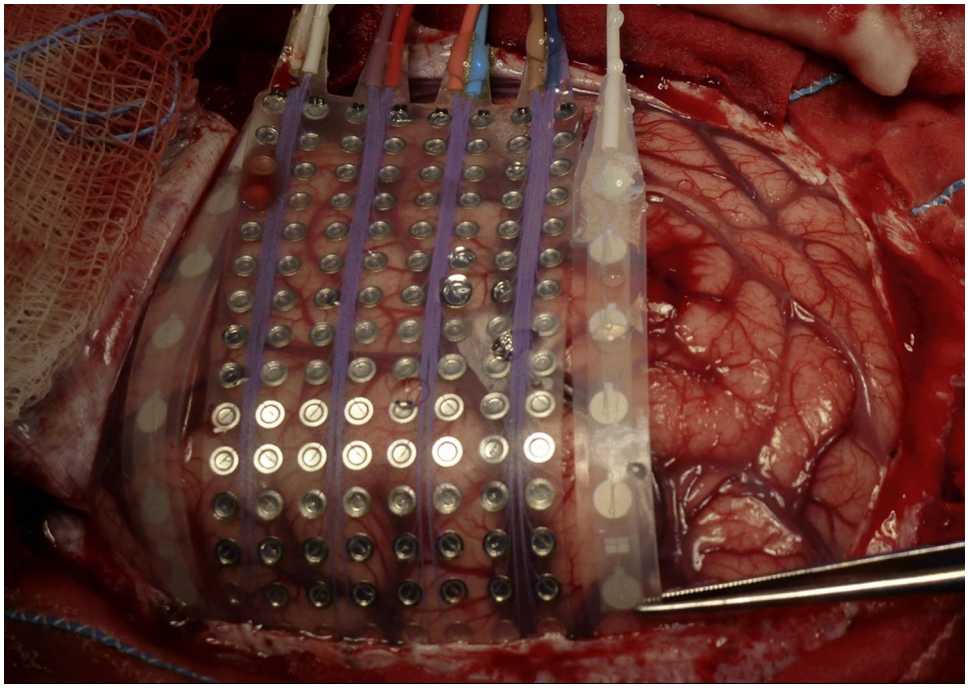

Fig. 1 High-density 128-channel (8 × 16) ECoG Grid. Photograph taken during subdural implantation. The electrodes are 2 mm in diameter and spaced 5 mm apart. Also visible in the figure are two 8 × 1 electrode strips with electrodes that are 4 mm in diameter and spaced 10 mm apart. Figure reused with permission from Ref. [33]

them minimally invasive. These electrodes can support broader spatial coverage but are limited in their density.

Electrocorticography (ECoG) uses 2D arrays of platinum-iridium disc electrodes embedded in soft silastic sheets that may be implanted in the subdural space to record EEG from the cortical surface (Fig. 1). The signals recorded with these electrodes are analogous to local field potentials (LFPs) recorded at larger spatial scales, which in turn depend on electrode size and spacing. ECoG recordings have been used extensively to identify the source of seizures in patients with drug-resistant epilepsy and to map cortical areas vital for brain function so that they may be preserved during resective surgery [15]. ECoG recordings in this patient population allowed the discovery of high gamma activity (~60–200 Hz) as a useful index of task-related local cortical activation [16], and subsequent studies in animals have shown that this activity is tightly coupled, both temporally and spatially, to changes in population firing rates in the immediate vicinity of recording electrodes [17, 18]. Indeed, differential changes in high gamma activity can be observed at electrodes separated by as little as 1 mm [19]. Thus, the surface area and spatial resolution of cortical representations that can be monitored with ECoG are limited only by the size and density of the electrode array used.

Target Population for Speech BCI

Because BCIs with adequate temporal and spatial resolution require surgically implanted electrodes, clinical trials of speech BCI devices are currently limited to patients with severe and permanent communication impairments, in whom the risk of surgical implantation can be justified by the severity of disability and a poor prognosis for recovery. The most pressing need for a speech BCI may be found in patients with LIS. Unlike patients who can rely on other means of communication, such as gestures and writing, LIS patients can typically only convey their thoughts through eye movements, eye blinking, or other minor residual movements. For patients with total locked in syndrome (TLIS) who have also lost the ability to control eye movement, this minimum means of communication is not even possible.

LIS is often caused by damage to the ventral pons, most commonly through an infarct, hemorrhage, or trauma, interrupting corticospinal tracts bilaterally and producing quadriplegia and anarthria [20, 21]. LIS can also be caused by degenerative neuromuscular diseases such as amyotrophic lateral sclerosis (ALS). In ALS, progressive weakness may result in LIS, especially if patients elect to have a tracheostomy and use artificial ventilation. Three categories of locked-in syndrome have been described: classic LIS where patients suffer from quadriplegia and anarthria but retain consciousness and vertical eye movement; incomplete LIS, in which patients have residual voluntary movement other than vertical eye movement; and TLIS, in which patients lose all motor function but remain fully conscious [22].

For LIS patients, anarthria arises from bilateral facio-glosso-pharyngo-laryngeal paralysis [23]. The cause of such paralysis in most LIS patients does not include speech-related cortical areas. Rather, anarthria reported in LIS patients usually results from interruption of neural pathways (corticobulbar tract) with loss of motor control of speech. Cranial nerve XII (hypoglossal nerves) controls the extrinsic muscles of the tongue: genioglossus, hyoglossus, styloglossus, and the intrinsic muscles of the tongue. These represent all muscles of the tongue except for the palatoglossus muscle [24]. Thus, lesions to neural pathways connecting cranial nerves XII produce a facial, tongue, and pharyngeal diplegia with anarthria, causing severe difficulties in swallowing and speech generation [25].

Another factor hindering speech function in LIS patients is impaired respiratory ability. Speech can be considered a sound exhalation and requires normal respiratory muscle strength. Normal speech requires active exhalation. Lesions of the ventral pons causing LIS not only impedes volitional behavior, but may also affect autonomous breathing [26].

Potential target populations for speech BCI also include patients suffering from aphasia. However, these patients often suffer from pathological changes in speech-related cortical regions, which would hinder the ability of a speech BCI to utilize natural speech circuitry for decoding [27]. While it is not impossible that the subject could be trained with a less natural neural control strategy, this extra challenge makes this population less suited for initial clinical trials.

Basic Principles of Operation

The underlying physiological support for a speech BCI is that distinct compositional features of speech can be represented by the weighted combinations of neural activity at subsets of recording electrodes [28]. Traditional BCI systems adopt techniques like linear discriminant analysis (LDA) to decode and classify speech into text before synthesizing audio through a conventional text-to-speech (TTS) application [29]. Recent studies have suggested the possibility of decoding neural signals directly using convolutional neural networks (CNN) to map high gamma activity recorded at different cortical sites onto speech features such as mel-spectrograms [30–32]. The decoded mel-spectrogram can then be used to recreate speech using a pre-trained neural network vocoder. The operation of a typical synthesis-based speech BCI is composed of four stages: recording of the raw neural signal, extraction of neural features from the raw signal, decoding of speech features from the neural features, and synthesis of audio from speech features (Fig. 2).

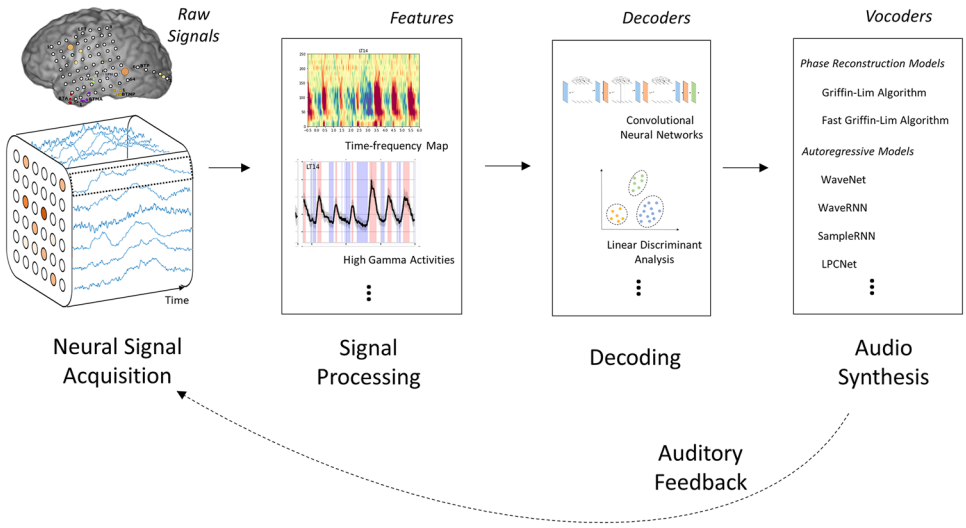

Fig. 2 Basic principles of operation of a speech BCI. During speech, raw neural signals are recorded and processed in real time. A decoder will then map processed neural signals into auditory features or textual transcriptions. Decoded features are then synthesized into audio waveforms and can potentially be played in real time as auditory feedback

Decoding of Speech-Related Neural Signals

The neurophysiological mechanisms responsible for speech production rely on semantic, auditory, and articulatory representations in cerebral cortex. Activation of these cortical representations can be measured and decoded individually (see detailed review in [33]). The decoding of semantic meaning [34, 35] and gestural representations [36–38] alone, however, does not translate into comprehensible speech without additional decoding linking them to linguistic features. Here, we consider only the aspects of speech that can be directly used in communication: from phoneme, to word, to sentence. Along with the grammar of a given language, these sub-units constitute the linguistic aspects of speech and directly support the textual decoding of speech neural signals. We will discuss how the acoustic representation, articulatory trajectories, or textual representations of these linguistic features can be mapped to neural signals (Fig. 3). However, linguistic features are not the only mediums that carry useful information in conversational speech. Paralinguistic features, such as pitch, tone, intonation, and prosody, convey important information and can significantly modify the meaning of speech. Therefore, we will also discuss speech synthesis which requires decoding both linguistic and paralinguistic aspects of speech.

Phoneme Decoding

Although decoding lower-level speech representations can potentially support the decoding of selected words with distinct semantic meaning, higher-level speech representation is preferred if the goal is to restore full conversational speech. One obvious candidate for decoding is the phoneme, the minimum distinguishable segment of speech. Early studies demonstrated the feasibility of phoneme-level neural decoding by classifying a limited set of vowels. Classification of 3 covertly articulated English vowels was achieved with up

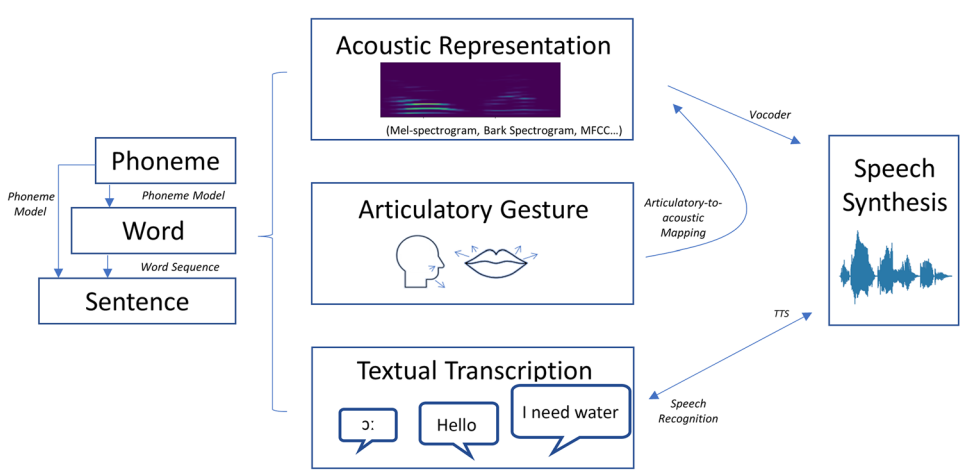

Fig. 3 Targets of speech neural signal decoding. The acoustic representation, articulatory trajectories, and textual representations of phonemes, words, or sentences are all potential targets for speech neural decoding

to 70% accuracy using spike data collected from an intracortically implanted microelectrode in a LIS patient [39]. Another study using a linear classifier trained on overt syllable speaking data collected from depth electrodes demonstrated 93% to near-perfect classification accuracy on 5 to 2 English vowels, respectively [40].

Similar classification on datasets consisting of a limited set of phonemes was also reported from ECoG studies. Blakely et al. [41] first demonstrated that phoneme pairs can be discriminated using ECoG data collected from a phoneme reading task. Pei et al. [42] classified 4 English vowels with Naïve Bayes classifiers trained on ECoG data collected in word repetition tasks, achieving 40.7% average classification accuracy for overt speech and 37.5% for covert speech. In the same study, they also showed above-chance decoding accuracy of four consonant pairs (leading and trailing consonants in a word). Ikeda et al. [43] adopted a linear classifier on 3 covertly articulated Japanese vowels, which were collected in an isolated vowel reading task. They were able to achieve a decoding accuracy of 42.2 to 46.7%. Apart from direct textual classification, linear classifiers were also used to decode acoustic formant features of 3 English vowels based on ECoG data collected from overt syllable reading [44]. Using spatiotemporal matched filters on ECoG data collected during overt isolated phoneme speaking tasks, Ramsey et al. [45] reported 75.5% decoding accuracy for 4 Dutch phonemes and rest. Milsap et al. [46] used similar spatiotemporal features in their neural voice activity detection study, successfully detecting all target keywords using neural templates trained from ECoG data from overt syllable reading tasks. Finally, one study also investigated the feasibility of classifying all English phonemes using ECoG data collected during overt word reading tasks, achieving 20.4% average decoding accuracy using LDA classifiers [47].

Word Decoding

Neural decoding may also target words, the smallest units of objective or practical meaning [48]. Relatively few studies have attempted to decode isolated words from neural data. For speech production studies, Kellis et al. [49] trained a linear classifier using ECoG data from overt word repetition tasks. They reported classification accuracy from 89.7 to 48% on vocabulary sizes from 2 to 10, respectively. Martin et al. [50] used support-vector machines for pairwise classification of words. They achieved 86.2% classification accuracy for overt speech production and 57.7% for covert speech using ECoG data recorded during repetition of isolated words. Apart from acoustic representation and textual classification, decoding of articulatory gestures from word-level speech neural data has also been investigated. Mugler et al. [38] used LDA and achieved 75% and 57.2% decoding accuracy (chance = 29.2% and 39.4%, respectively) for 13 articulator types in two subjects.

Sentence Decoding

As a self-contained and complete vehicle for speech, sentence is the core unit of language interpretation [51]. Several advantages come with decoding whole sentences: (1) It is a more natural paradigm for communication; (2) The broader spatial distribution of sentence-level speech and richer temporal information could offer more information for decoding. (3) The incorporation of language models can increase the decoding accuracy. Recent years have seen the growing popularity of sentence-level decoding. Martin et al. [28] successfully reconstructed spectro-temporal features of sentence-level speech from ECoG recordings of both overt and covert sentence reading. Moses et al. [52] used LDA to classify sentence-level speech perception data in real time. They proposed both a direct classification approach where sentence-level neural activities were used to train the decoder and a continuous phoneme classification approach similar to their previous method in [53]. Herff et al. [54] developed one of the first functioning systems transcribing neural activities during overt sentence production into textual output. They used Bayesian update [55] to combine a statistical ECoG phone model with a language model and predict the most likely sequence of words. The word error rate (WER) of this system ranged between 60 and 15% for vocabulary sizes between 100 and 10, respectively. More recently, deep learning architectures for automatic speech recognition (ASR) were used in sentence-level neural decoding, significantly improving decoding accuracy. Makin et al. [56] used Encoder-Decoder recurrent neural networks (RNNs, [57]) to make sequence-to-sequence predictions. In contrast to other studies mentioned earlier in this section, this study mapped neural activities recorded from ECoG grids into word sequences instead of phoneme sequences. They achieved a 3% WER for a single participant with a vocabulary size of about 250. In another end-to-end sentence decoding study, Sun et al. [58] proposed a deep learning architecture consisting of a neural feature encoder network trained to extract spatiotemporal neural features, feature regularization networks trained to force meaningful representation in latent space, and a text decoder network trained to minimize alignment-free connectionist temporal classification loss. With a language model, their study achieved a WER of 10.6%, 8.5%, and 7.0% on three different subjects with vocabulary sizes from 1200 to 1900. Recently, Moses et al. [59] successfully achieved online sentence decoding using chronically implanted 128-channel ECoG grid. Training data was collected during attempted unintelligible overt speech from a participant with quadriplegia and anarthria resulting from a pontine stroke. To decode sentences, they

first trained a neural network to detect the individual word in speech. Subsequent neural networks were trained to classify detected words into one of the 50 words from the limited vocabulary set used in the study. The accuracy of the classification model in offline analysis was 47.1% (chance accuracy was 2%). Two additional models were used in sentence decoding. The first was a language model that predicted the probability that a word would occur in the English language given the sequence of words preceding it. This model was trained on a custom dataset consisting sentence sequences constructed with words from the aforementioned 50-word set. The second was a Viterbi decoder that combined the probability from the language model and the word classification model to make a final prediction. The study achieved real-time sentence decoding with a median WER of 25.6% (chance WER was 92.1%). The median number of words decoded per minute was 15.2. Overall, these studies demonstrated the feasibility of transcribing neural data into textual output.

Speech Synthesis

One of the challenges in developing classification-based decoding methods is the variability of speech. Even for a single speaker, speech signals are impacted by the rate of speech, coarticulation, emotional state, and vocal effort [60, 61]. To produce textual output, decoding models need to be robust to these variabilities. At the same time, some of these variabilities carry linguistic meaning and constitute an essential part of natural speech. For example, prosody and intonation are often used for conveying humorous or satirical intents, as are pauses and varying rates of speech for emphasis. By directly mapping speech neural signals onto acoustic speech or speech-related features, researchers have been able to preserve these non-representational and paralinguistic aspects of natural speech. Herff et al. [62] proposed a method to improve on previous classification studies. They used a pattern matching approach for neural activities and concatenated the corresponding ground-truth speech units to generate continuous audio. Their unit-selection model was trained on small sets of ECoG data (8.3 to 11.7 min) and simultaneous audio recordings during overt speaking tasks. This study demonstrated that intelligible speech could be generated using models that were less demanding on computing resources and that were trained on limited sets of data.

The use of deep learning models significantly improved the performance of synthesis-based speech BCIs. Using data recorded from ECoG grids and stereo-electroencephalographic (sEEG) depth arrays during speech perception, Akbari et al. [63] showed intelligible synthesis of sentences and isolated digits using a standard feedforward network mapping ECoG high gamma, as well as low-frequency signal features, to vocoder parameters, including spectral envelope, pitch, voicing, and aperiodicity. They achieved a 65% relative increase in intelligibility over a baseline linear regression model.

Recently, studies based on deep learning methods also demonstrated the feasibility of synthesis from speech production data. Angrick et al. [30] showed that high-quality audio of overtly spoken words could be reconstructed from ECoG recordings using two consecutive deep neural networks (DNNs). Their first DNN consisted of densely connected neural networks [64] and mapped neural features into spectral acoustic representations. These speech representations were then reconstructed into audio waveforms by WaveNet [65], a secondary vocoder DNN. Anumanchipalli et al. [32] reconstructed spoken sentences from ECoG data using two recurrent bidirectional long-term short-term memory networks (bLSTM) [66]. Their first bLSTM mapped ECoG high gamma activity onto vocal tract trajectories (inferred statistically) from speaking full sentences. The second bLSTM then inverted the trajectories to acoustic speech features. Finally, an HMM-based excitation model synthesized speech waveforms based on these speech features [67]. They also showed that their network generalized to unseen sentences and to silently mouthed speech without vocalization. In both studies, neural activities were mapped first into intermediary speech representations, from which speech waveforms were subsequently reconstructed. Both studies showed reasonable speech reconstruction using relatively small amounts of data. Angrick et al. [30] used datasets between 8.3 and 11.7 min, and Anumanchipalli et al. [32] showed robust decoding performance with a minimum of 25 min of data. The fact that both studies were able to achieve intelligible speech synthesis with limited data size with the incorporation of intermediary speech representation might point to the particular usefulness of leveraging speech-adjacent features to train models in data-limited settings. A recent study by Kohler et al. [31] also examined the possibility of using an encoder-decoder sequence-to-sequence model to predict spectral acoustic representation from sEEG signals collected during overt speech. Audio waveforms were then reconstructed using a WaveGlow [68] vocoder. Together, these findings demonstrate the strong potential of neural networks in decoding and synthesizing speech neural data. These studies also suggest the benefits of having consecutive neural networks with distinct roles bridging neural activities, intermediary speech representations, and eventually auditory speech reconstruction.

Vocoders for Speech Synthesis from Neural Signals

One key component of synthesis-based speech BCI is the vocoder, which generates a natural-sounding human voice either from textual representations or acoustic features,

depending on the targets for neural decoding. A text-to-speech system (TTS) is often used to synthesize speech acoustic waveforms from word or sentence-level textual input, making it suitable for providing auditory feedback after textual transcription has been decoded from neural signals. Early TTS systems relied on unit-selection approaches, concatenating small segments of speech to generate continuous waveforms [69, 70]. Herff et al. [62] took the same unit-selection approach for speech synthesis in their neural decoding study but bypassed the intermediate text representation. Subsequently, statistical parameter speech synthesis (SPSS) grew in popularity. SPSS models map linguistic features from text into intermediate acoustic features and reconstruct speech waveforms from these features [71]. Supplied with textual input from neural decoders, these models have been used to generate audio outputs.

Crucially, SPSS models can be used not only in the vocoding of textual output, but also acoustic features decoded from neural signals. Provided the same acoustic representation is used in training the neural decoder and SPSS model, acoustic waveforms can be reconstructed directly from the vocoder component of the SPSS models. Vocoders used in SPSS can generally be divided into two categories, autoregressive (AR) probabilistic models and phase estimation models. Both have benefited from incorporating deep learning techniques. For phase estimation vocoders, the classic Griffin-Lim algorithm (GLA) is still one of the most used. GLA inverts the spectrogram based on the redundancy of short-time Fourier transformation [72]. A faster algorithm inspired by GLA has also been proposed, improving both the quality and the speed of the original algorithm [73]. A GLA vocoder was used in one neural decoding study to reconstruct speech waveforms from quantized spectrograms predicted from sentence-level ECoG data [74].

In contrast to phase estimation models, classical AR models have attempted to synthesize speech by finding parameters for the source-filter model (see detailed review in [75]). An HMM-based excitation model [67] was used by Anumanchipalli et al. [32] to reconstruct acoustic waveforms from intermediate acoustic features predicted from their neural decoder. In recent years, several deep-learning–based AR vocoders have also shown great promises for use in speech BCI, including WaveNet [65], SampleRNN [76], WaveRNN [77], and LPCNet [78]. WaveNet was used in one neural decoding study to reconstruct auditory waveforms from decoded acoustic representations [30] and one speech synthesis study based on electromyography [79]. WaveGlow, a flow-based method without the need for autoregression inspired by Glow and WaveNet, has also been proposed [68]. Recently used by Kohler et al. [31] for offline synthesis in their sEEG-based speech decoding system, it could be a promising candidate as a vocoder in a real-time speech BCI system due to its fast inference time.

Challenges and Future Directions

Chronic ECoG

Most of the speech BCI studies we reviewed above have been based on acute or short-term ECoG recordings for clinical purposes, mostly for surgical treatment of drug-resistant epilepsy, but also brain tumors [15, 80]. Long-term ECoG signal stability for speech decoding has not yet been fully investigated. However, motor BCI research based on long-term ECoG signals have demonstrated reliable decoding from chronic implants [81, 82]. The safety and stability of ECoG implants in individuals with late-stage ALS have also been reported. For over 36 months, the motor-based system maintained high performance and was increasingly utilized by the study participant [4, 83]. In one study using the NeuroPace RNS System with sparse electrode coverage, long-term stability of speech-evoked cortical responses was observed [84]. A recently published study examined the feasibility of speech decoding using a chronically implanted 128-channel ECoG grid. The study lasted 81 weeks with 50 experimental sessions conducted at the participant's home and a nearby office. The authors reported that the ECoG signals collected for this study were stable across the study period for decoding purposes [59]. Beyond the aforementioned studies, the safety of long-term ECoG implantation has been established by multiple studies in non-human primates [85, 86]. These studies indicate that a chronic ECoG implant for speech BCI should be safe and should provide stable signal quality.

Real-time Speech Decoding and Synthesis

Assistive speech BCI systems for patients with LIS need to operate in real time with reasonably low latency. For systems designed to provide a classification-based selection or textual transcription, the latency can be longer at the expense of the information transfer rate [87]. Studies have shown that a real-time ECoG classification system is indeed feasible for sentence-level speech perception [52] and overt phrase/word-level speech production [88]. The drawback of such system is the lack of immediate auditory feedback, which plays an important role in the speech production process [89], and the lack of other expressive features of spoken acoustics.

For patients with LIS, a speech BCI system capable of providing real-time auditory feedback could be very useful. Timely sensory feedback, though artificial, can allow users to make adjustments in vocoding efforts and to detect and correct errors. Although individuals retain the ability to produce intelligible speech years after loss of hearing,

their speech deteriorates over time due to the lack of feedback [90–92]. Even though most LIS patients retain intact hearing [21], the same deterioration of speaking abilities might occur due to the absence of self-generated speech, and consequently, feedback from it. More importantly, since speaking with a synthesis-based BCI system is significantly different from speaking prior to loss of function, recalibration or even relearning of speech production is needed, thus requiring real-time auditory feedback [93, 94]. Although no ECoG-based online speech synthesis has yet been reported, several studies have explored closed-loop speech synthesis using neurotrophic electrodes [39], stereo-electroencephalography [95], and electromyography [96], to varying degrees of intelligibility.

For synthesis-based speech BCIs aiming to provide auditory feedback, latency must be kept at a minimum to avoid disruption of speech production. Previous evidence suggests that acoustic feedback at a 200 ms latency can disrupt adult speech production [97]. Although slow and prolonged speech can be maintained at longer delays than 200 ms, shorter delays are needed in fast-paced natural speech [98, 99]. Studies in delayed auditory feedback have found that delays less than 75 ms are hardly perceptible to speakers, and fast-paced speech can be maintained with such delay, while optimal delay is less than 50 ms [100–102].

Decoding Silent Speech

Many of the studies we reviewed here were based on overt speech production in which subjects clearly enunciated their speech and produced normal acoustic speech waveforms. This acoustic output can be critically useful for training speech decoders and for providing ground truth when attempting to segment neural signals that correspond with spoken words, phrases, or sentences. However, for patients who are locked-in, overt speech production is severely impacted, if not outright impossible. Therefore, speech BCI systems for patients with LIS may need to be trained on and decode silent speech. Speech can be silent either because no attempt is made to phonate or articulate (covert speech) or because articulation occurs without phonation (mimed speech). In patients with different degrees of paralysis of the muscles for phonation and articulation, speech may be silent even though the patient is attempting to phonate and/or articulate (attempted speech). In overt speech studies, training labels are easily obtainable during neural recording sessions in the form of simultaneous audio recording. For silent speech, experimental paradigms need to be carefully designed for subjects to vocalize with predictable and precise timing. Such experiments are even more challenging with LIS patients, who have difficulty in giving feedback, verbally or otherwise.

Compared to decoding overt speech, silent speech not only fails to provide a ground truth for training but may also produce different patterns of cortical activation. Indeed, most studies of covert speech have shown that it is accompanied by far less cortical activation than overt speech. Moreover, the cortical representations of covert speech may differ from those of overt speech, making it more difficult to adapt successful decoding methods from overt studies to use in LIS patients [103, 104]. Despite these challenges, multiple studies have shown success in phoneme [42, 43], word [50], and sentence classification [28] from ECoG signals (see detailed review of covert speech decoding in [105]). Moreover, in patients with paralysis of speech musculature, cortical activation during attempted speech is comparable to that observed during overt speech in able normal subjects [106]. In addition, progress has been made in synthesizing speech from silently articulated speech (mimed speech) in which subjects move articulators without vocalization [32]. A closed-loop online speech synthesis system based on covert speech has also been proposed [95]. However, online speech synthesis with reasonable intelligibility from silent speech has not yet been achieved at the time of this review.

Conclusions

This review summarizes previous studies on speech decoding from ECoG signals in the larger context of BCI as an alternative and augmentative channel for communication. Different levels of speech representations: phonemes, words, and sentences may be classified from neural signals. Emerging interest in adopting deep learning in neural speech decoding has yielded promising results. Breakthroughs have also been made in directly synthesizing spoken acoustics from ECoG recordings. We also discuss several challenges that must be overcome in developing a synthesis-based speech BCI for patients with LIS, such as the need for a safe and effective chronically implanted ECoG array with sufficient density and coverage of cortical speech areas, and a real-time system capable of decoding covert or attempted speech in the absence of acoustic output. Despite these challenges, progress continues to advance toward providing an alternate method of speaking for patients with LIS and other severe communication disorders.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13311-022-01190-2.

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

Funding The authors are supported by the National Institutes of Health under award number UH3NS114439 (NINDS) and U01DC016686 (NIDCD).

References

1. Brumberg JS, Nieto-Castanon A, Kennedy PR, et al. Brain-computer interfaces for speech communication. Speech Commun. 2010:367–79.
2. Wolpaw JR, Birbaumer N, McFarland DJ, et al. Brain-computer interfaces for communication and control. Clin Neurophysiol. 2002:767–91.
3. Kübler A, Furdea A, Halder S, et al. A brain–computer interface controlled auditory event-related potential (P300) spelling system for locked-in patients. Ann N Y Acad Sci. 2009;1157:90–100.
4. Vansteensel M, Pels E, Bleichner M, et al. Fully implanted brain-computer interface in a locked-in patient with ALS. N Engl J Med. 2016:2060–66.
5. Willett FR, Avansino DT, Hochberg LR, et al. High-performance brain-to-text communication via handwriting. Nature. 2021;593:249–54.
6. Lesenfants D, Habbal D, Lugo Z, et al. An independent SSVEP-based brain–computer interface in locked-in syndrome. J Neural Eng. 2014;11:035002.
7. Chang EF, Anumanchipalli GK. Toward a speech neuroprosthesis. JAMA. 2020;323:413–4.
8. Pandarinath C, Nuyujukian P, Blabe CH, et al. High performance communication by people with paralysis using an intracortical brain-computer interface. Kastner S, editor. Elife. 2017;6:e18554.
9. Dassios G, Fokas A, Kariotou F. On the non-uniqueness of the inverse MEG problem. Inverse Probl. 2005:L1–L5.
10. Im C, Seo J-M. A review of electrodes for the electrical brain signal recording. Biomed Eng Lett. 2016:104–12.
11. Musk E, Neuralink. An integrated brain-machine interface platform with thousands of channels. J Med Internet Res. 2019;21:e16194.
12. Wilson GH, Stavisky SD, Willett FR, et al. Decoding spoken English from intracortical electrode arrays in dorsal precentral gyrus. J Neural Eng. 2020;17:066007.
13. Stavisky SD, Willett FR, Wilson GH, et al. Neural ensemble dynamics in dorsal motor cortex during speech in people with paralysis. Makin TR, Shinn-Cunningham BG, Makin TR, et al., editors. Elife. 2019;8:e46015.
14. Herff C, Krusienski DJ, Kubben P. The potential of stereotactic-eeg for brain-computer interfaces: current progress and future directions. Front Neurosci. 2020;14:123.
15. Crone NE, Sinai A, Korzeniewska A. High-frequency gamma oscillations and human brain mapping with electrocorticography. In: Neuper C, Klimesch W, editors. Prog Brain Res [Internet]. Elsevier; 2006 [cited 2021 May 31]. p. 275–295. Available from: https://www.sciencedirect.com/science/article/pii/S0079612306590193.
16. Crone NE, Korzeniewska A, Franaszczuk PJ. Cortical gamma responses: searching high and low. Int J Psychophysiol. 2011;79:9–15.
17. Ray S, Crone NE, Niebur E, et al. Neural correlates of high-gamma oscillations (60–200 Hz) in Macaque local field potentials and their potential implications in electrocorticography. J Neurosci. 2008;28:11526–36.
18. Ray S, Hsiao SS, Crone NE, et al. Effect of stimulus intensity on the spike–local field potential relationship in the secondary somatosensory cortex. J Neurosci. 2008;28:7334–43.
19. Slutzky MW, Jordan LR, Krieg T, et al. Optimal spacing of surface electrode arrays for brain–machine interface applications. J Neural Eng. 2010;7:026004.
20. León-Carrión J, Eeckhout PV, Domínguez-Morales MDR. Review of subject: the locked-in syndrome: a syndrome looking for a therapy. Brain Inj. 2002:555–69.
21. Smith E, Delargy M. Locked-in syndrome. Bmj. 2005:406–09.
22. Bauer G, Gerstenbrand F, Rumpl E. Varieties of the locked-in syndrome. J Neurol. 1979:77–91.
23. Richard I, Péreon Y, Guiheneu P, et al. Persistence of distal motor control in the locked in syndrome. Review of 11 patients. Paraplegia. 1995:640–46.
24. Mtui E, Gruener G, Dockery P, et al. Fitzgerald's clinical neuroanatomy and neuroscience. Edition 7. Philadelphia, PA: Elsevier; 2017.
25. Leon-Carrion J, von Wild KRH, Zitnay GA. Brain injury treatment: theories and practices. Taylor & Francis; 2006.
26. Heywood P, Murphy K, Corfield D, et al. Control of breathing in man; insights from the "locked-in" syndrome. Respir Physiol. 1996:13–20.
27. Gorno-Tempini M, Hillis A, Weintraub S, et al. Classification of primary progressive aphasia and its variants. Neurology. 2011:1006–14.
28. Martin S, Brunner P, Holdgraf C, et al. Decoding spectrotemporal features of overt and covert speech from the human cortex. Front Neuroeng. 2014:14.
29. Soman S, Murthy B. Using brain computer interface for synthesized speech communication for the physically disabled. Procedia Comput Sci. 2015:292–98.
30. Angrick M, Herff C, Mugler E, et al. Speech synthesis from ECoG using densely connected 3D convolutional neural networks. J Neural Eng. 2019.
31. Kohler J, Ottenhoff MC, Goulis S, et al. Synthesizing speech from intracranial depth electrodes using an encoder-decoder framework. ArXiv211101457 Cs [Internet]. 2021 [cited 2022 Jan 3]; Available from: http://arxiv.org/abs/2111.01457.
32. Anumanchipalli G, Chartier J, Chang E. Speech synthesis from neural decoding of spoken sentences. Nature. 2019:493.
33. Rabbani Q, Milsap G, Crone NE. The potential for a speech brain-computer interface using chronic electrocorticography. Neurotherapeutics. 2019:144–65.
34. Chen Y, Shimotake A, Matsumoto R, 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.
35. Rupp K, Roos M, Milsap G, et al. Semantic attributes are encoded in human electrocorticographic signals during visual object recognition. Neuroimage. 2017:318–29.
36. Lotte F, Brumberg JS, Brunner P, et al. Electrocorticographic representations of segmental features in continuous speech. Front Hum Neurosci [Internet]. 2015 [cited 2021 May 18];9. Available from: https://www.frontiersin.org/articles/10.3389/fnhum.2015.00097/full.
37. Mugler EM, Tate MC, Livescu K, et al. Differential representation of articulatory gestures and phonemes in precentral and inferior frontal gyri. J Neurosci. 2018;38:9803–13.
38. Mugler EM, Goldrick M, Rosenow JM, et al. Decoding of articulatory gestures during word production using speech motor and premotor cortical activity. 2015 37th Annu Int Conf IEEE Eng Med Biol Soc EMBC. 2015. p. 5339–5342.
39. Guenther F, Brumberg J, Wright E, et al. A wireless brain-machine interface for real-time speech synthesis. PLoS One. 2009.
40. Tankus A, Fried I, Shoham S. Structured neuronal encoding and decoding of human speech features. Nat Commun. 2012;3:1015.
41. Blakely T, Miller KJ, Rao RPN, et al. Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. 2008 30th Annu Int Conf IEEE Eng Med Biol Soc. 2008. p. 4964–4967.
42. Pei X, Barbour DL, Leuthardt EC, et al. Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. J Neural Eng. 2011;8:046028.
43. Ikeda S, Shibata T, Nakano N, et al. Neural decoding of single vowels during covert articulation using electrocorticography. Front Hum Neurosci [Internet]. 2014 [cited 2021 May 14];8.

Available from: https://www.frontiersin.org/articles/10.3389/fnhum.2014.00125/full.
44. Bouchard KE, Chang EF. Neural decoding of spoken vowels from human sensory-motor cortex with high-density electrocorticography. 2014 36th Annu Int Conf IEEE Eng Med Biol Soc. 2014. p. 6782–6785.
45. Ramsey NF, Salari E, Aarnoutse EJ, et al. Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. Neuroimage. 2018;180:301–11.
46. Milsap G, Collard M, Coogan C, et al. Keyword spotting using human electrocorticographic recordings. Front Neurosci. 2019.
47. Mugler E, Patton J, Flint R, et al. Direct classification of all American English phonemes using signals from functional speech motor cortex. J Neural Eng. 2014.
48. Sapir E. Language: an introduction to the study of speech. Brace: Harcourt; 1921.
49. Kellis S, Miller K, Thomson K, et al. Decoding spoken words using local field potentials recorded from the cortical surface. J Neural Eng. 2010;7:056007.
50. Martin S, Brunner P, Iturrate I, et al. Word pair classification during imagined speech using direct brain recordings. Sci Rep. 2016;6:25803.
51. Chomsky N. Syntactic structures [Internet]. Syntactic Struct. De Gruyter Mouton; 2009 [cited 2021 Oct 21]. Available from: https://www.degruyter.com/document/doi/10.1515/9783110218329/html.
52. Moses D, Leonard M, Chang E. Real-time classification of auditory sentences using evoked cortical activity in humans. J Neural Eng. 2018.
53. Moses D, Mesgarani N, Leonard M, et al. Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. J Neural Eng. 2016.
54. Herff C, Heger D, de Pesters A, et al. Brain-to-text: decoding spoken phrases from phone representations in the brain. Front Neurosci. 2015.
55. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989;77:257–86.
56. Makin JG, Moses DA, Chang EF. Machine translation of cortical activity to text with an encoder–decoder framework. Nat Neurosci. 2020:575–82.
57. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. Adv Neural Inf Process Syst [Internet]. Curran Associates, Inc.; 2014 [cited 2021 Dec 16]. Available from: https://proceedings.neurips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html.
58. Sun P, Anumanchipalli GK, Chang EF. Brain2Char: a deep architecture for decoding text from brain recordings. J Neural Eng. 2020;17:066015.
59. Moses DA, Metzger SL, Liu JR, et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. N Engl J Med. 2021;385:217–27.
60. Benzeghiba M, De Mori R, Deroo O, et al. Automatic speech recognition and speech variability: a review. Speech Commun. 2007;49:763–86.
61. Zelinka P, Sigmund M, Schimmel J. Impact of vocal effort variability on automatic speech recognition. Speech Commun. 2012;54:732–42.
62. Herff C, Diener L, Angrick M, et al. Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices. Front Neurosci. 2019.
63. Akbari H, Khalighinejad B, Herrero J, et al. Towards reconstructing intelligible speech from the human auditory cortex. Sci Rep. 2019.
64. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. 2017 IEEE Conf Comput Vis Pattern Recognit CVPR. 2017:2261–69.
65. Oord A van den, Dieleman S, Zen H, et al. Wavenet: a generative model for raw audio. ArXiv Prepr ArXiv160903499. 2016.
66. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. 2005;18:602–10.
67. Maia R, Toda T, Zen H, et al. A trainable excitation model for HMM-based speech synthesis. Eighth Annu Conf Int Speech Commun Assoc. 2007.
68. Prenger R, Valle R, Catanzaro B. Waveglow: a flow-based generative network for speech synthesis. ICASSP 2019 - 2019 IEEE Int Conf Acoust Speech Signal Process ICASSP. 2019:3617–21.
69. Black AW, Taylor PA. Automatically clustering similar units for unit selection in speech synthesis. International Speech Communication Association; 1997 [cited 2021 May 23]. Available from: https://era.ed.ac.uk/handle/1842/1236.
70. Hunt AJ, Black AW. Unit selection in a concatenative speech synthesis system using a large speech database. 1996 IEEE Int Conf Acoust Speech Signal Process Conf Proc. 1996. p. 373–376 vol. 1.
71. Wang X, Lorenzo-Trueba J, Takaki S, et al. A comparison of recent waveform generation and acoustic modeling methods for neural-network-based speech synthesis. 2018 IEEE Int Conf Acoust Speech Signal Process ICASSP. IEEE; 2018. p. 4804–4808.
72. Griffin D, Lim J. Signal estimation from modified short-time Fourier transform. ICASSP 83 IEEE Int Conf Acoust Speech Signal Process. 1983. p. 804–807.
73. Perraudin N, Balazs P, Søndergaard PL. A fast Griffin-Lim algorithm. 2013 IEEE Workshop Appl Signal Process Audio Acoust. 2013. p. 1–4.
74. Angrick M, Herff C, Johnson G, et al. Speech spectrogram estimation from intracranial brain activity using a quantization approach. Interspeech 2020 [Internet]. ISCA; 2020 [cited 2021 May 23]. p. 2777–2781. Available from: http://www.isca-speech.org/archive/Interspeech_2020/abstracts/2946.html.
75. Airaksinen M, Juvela L, Bollepalli B, et al. A comparison between STRAIGHT, glottal, and sinusoidal vocoding in statistical parametric speech synthesis. IEEEACM Trans Audio Speech Lang Process. 2018;26:1658–70.
76. Mehri S, Kumar K, Gulrajani I, et al. SampleRNN: an unconditional end-to-end neural audio generation model. ArXiv Prepr ArXiv161207837. 2016.
77. Kalchbrenner N, Elsen E, Simonyan K, et al. Efficient neural audio synthesis. Proc 35th Int Conf Mach Learn [Internet]. PMLR; 2018 [cited 2021 Dec 16]. p. 2410–2419. Available from: https://proceedings.mlr.press/v80/kalchbrenner18a.html.
78. Valin J, Skoglund J. LPCNET: improving neural speech synthesis through linear prediction. ICASSP 2019 - 2019 IEEE Int Conf Acoust Speech Signal Process ICASSP. 2019. p. 5891–5895.
79. Gaddy D, Klein D. Digital Voicing of Silent Speech. Proc 2020 Conf Empir Methods Nat Lang Process EMNLP [Internet]. Online: Association for Computational Linguistics; 2020 [cited 2022 Jan 3]. p. 5521–5530. Available from: https://aclanthology.org/2020.emnlp-main.445.
80. Caldwell DJ, Ojemann JG, Rao RPN. Direct electrical stimulation in electrocorticographic brain-computer interfaces: enabling technologies for input to cortex. Front Neurosci. 2019:804.
81. Benabid AL, Costecalde T, Eliseyev A, et al. An exoskeleton controlled by an epidural wireless brain–machine interface in a tetraplegic patient: a proof-of-concept demonstration. Lancet Neurol. 2019;18:1112–22.
82. Silversmith DB, Abiri R, Hardy NF, et al. Plug-and-play control of a brain–computer interface through neural map stabilization. Nat Biotechnol. 2021;39:326–35.
83. Pels EGM, Aarnoutse EJ, Leinders S, et al. Stability of a chronic implanted brain-computer interface in late-stage amyotrophic lateral sclerosis. Clin Neurophysiol. 2019;130:1798–803.

84. Rao VR, Leonard MK, Kleen JK, et al. Chronic ambulatory electrocorticography from human speech cortex. Neuroimage. 2017;153:273–82.
85. Chao ZC, Nagasaka Y, Fujii N. Long-term asynchronous decoding of arm motion using electrocorticographic signals in monkey. Front Neuroengineering [Internet]. 2010 [cited 2021 May 31];3. Available from: https://www.frontiersin.org/articles/10.3389/fneng.2010.00003/full.
86. Degenhart A, Eles J, Dum R, et al. Histological evaluation of a chronically-implanted electrocorticographic electrode grid in a non-human primate. J. Neural Eng. 2016.
87. Chesters J, Baghai-Ravary L, Möttönen R. The effects of delayed auditory and visual feedback on speech production. J Acoust Soc Am. 2015;137:873–83.
88. Moses D, Leonard MK, Makin JG, et al. Real-time decoding of question-and-answer speech dialogue using human cortical activity. Nat Commun. 2019;10:3096.
89. Guenther FH, Hickok G. Role of the auditory system in speech production. Handb Clin Neurol. 2015;129:161–75.
90. Cowie R, Douglas-Cowie E, Kerr AG. A study of speech deterioration in post-lingually deafened adults. J Laryngol Otol. 1982;96:101–12.
91. Perkell JS, Lane H, Denny M, et al. Time course of speech changes in response to unanticipated short-term changes in hearing state. J Acoust Soc Am. 2007;121:2296–311.
92. Waldstein RS. Effects of postlingual deafness on speech production: implications for the role of auditory feedback. J Acoust Soc Am. 1990;88:2099–114.
93. Kent RD. Research on speech motor control and its disorders: a review and prospective. J Commun Disord. 2000:391–428.
94. Perkell JS, Guenther FH, Lane H, et al. A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. J Phon. 2000;28:233–72.
95. Angrick M, Ottenhoff MC, Diener L, et al. Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. Commun Biol. 2021;4:1–10.
96. Bocquelet F, Hueber T, Girin L, et al. Real-time control of an articulatory-based speech synthesizer for brain computer interfaces. PLoS Comput Biol. 2016;12:e1005119.
97. MacKay DG. Metamorphosis of a critical interval: age-linked changes in the delay in auditory feedback that produces maximal disruption of speech. J Acoust Soc Am. 1968;43:811–21.
98. Antipova EA, Purdy SC, Blakeley M, et al. Effects of altered auditory feedback (AAF) on stuttering frequency during monologue speech production. J Fluen Disord. 2008;33:274–90.
99. Lincoln M, Packman A, Onslow M. Altered auditory feedback and the treatment of stuttering: a review. J Fluen Disord. 2006;31:71–89.
100. Kalinowski J, Stuart A. Stuttering amelioration at various auditory feedback delays and speech rates. Eur J Disord Commun J Coll Speech Lang Ther Lond. 1996;31:259–69.
101. Stuart A, Kalinowski J, Rastatter MP, et al. Effect of delayed auditory feedback on normal speakers at two speech rates. J Acoust Soc Am. 2002;111:2237–41.
102. Zimmerman S, Kalinowski J, Stuart A, et al. Effect of altered auditory feedback on people who stutter during scripted telephone conversations. J Speech Lang Hear Res. 1997;40:1130–4.
103. Proix T, Saa JD, Christen A, et al. Imagined speech can be decoded from low- and cross-frequency features in perceptual space. bioRxiv. 2021. https://doi.org/10.1101/2021.01.26.428315.
104. Tian X, Poeppel D. Mental imagery of speech and movement implicates the dynamics of internal forward models. Front Psychol [Internet]. 2010 [cited 2021 May 31];1. Available from: https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00166/full.
105. Martin S, Iturrate I, Millan J, et al. Decoding inner speech using electrocorticography: progress and challenges toward a speech prosthesis. Front Neurosci. 2018.
106. Bleichner MG, Jansma JM, Salari E, et al. Classification of mouth movements using 7 T fMRI. J Neural Eng. 2015;12:066026.

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