



Published in final edited form as:

Expert Rev Med Devices. 2010 September ; 7(5): 667–679. doi:10.1586/erd.10.34.

Development of speech prostheses: current status and recent advances

Jonathan S Brumberg^{1,†} and Frank H Guenther^{1,2,3,4}

¹Department of Cognitive and Neural Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA

²Department of Speech, Language and Hearing Sciences, College of Health and Rehabilitation Sciences: Sargent College, Boston University, Boston, MA 02215, USA

³Division of Health Sciences and Technology, Harvard University – Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁴Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Charlestown, MA 02129, USA

Abstract

Brain–computer interfaces (BCIs) have been developed over the past decade to restore communication to persons with severe paralysis. In the most severe cases of paralysis, known as locked-in syndrome, patients retain cognition and sensation, but are capable of only slight voluntary eye movements. For these patients, no standard communication method is available, although some can use BCIs to communicate by selecting letters or words on a computer. Recent research has sought to improve on existing techniques by using BCIs to create a direct prediction of speech utterances rather than to simply control a spelling device. Such methods are the first steps towards speech prostheses as they are intended to entirely replace the vocal apparatus of paralyzed users. This article outlines many well known methods for restoration of communication by BCI and illustrates the difference between spelling devices and direct speech prediction or speech prosthesis.

Keywords

brain–computer interface; speech prosthesis; speech synthesis

The past decade has seen a rapid proliferation of brain–computer interface (BCI) research, particularly for communication. The primary target populations for such BCI applications are persons with severe paralysis, who can only communicate using computer interfaces designed to read and interpret neurological signals. For example, patients with amyotrophic lateral sclerosis (ALS) or locked-in syndrome (LIS) [1] are often cited as ideal users of BCI technologies. ALS is a chronic, progressive neurological disorder in which motor neurons in both the patients brain and spinal cord degenerate, reducing their ability to properly actuate the peripheral nervous and muscular systems, including the vocal apparatus. The disease results

© 2010 Expert Reviews Ltd

[†]Author for correspondence: Tel.: +1 617 353 6181, Fax: +1 617 353 7755, brumberg@bu.edu.

Financial & competing interests disclosure

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

in progressively worsening paralysis, eventually rendering patients anarthric. LIS describes the state of total or near-total paralysis with intact sensation and cognition; LIS patients may retain some slight ocular or facial movements. Late-stage ALS often results in LIS, although other causes exist, most notably brainstem stroke. Persons with LIS but no neurodegenerative complications can have extended life expectancies with 5- and 10-year survival rates over 80% for those surviving the first year [2–5], and any means of social interaction can greatly improve their quality of life. In this article, we will discuss the options, in particular for direct speech production.

Generally, BCI research aims to provide a direct link between neural activity and external devices, such as mouse cursors or robotic limbs. Thus, BCI communication systems often involve an intermediate step between brain activity and speech or verbal output. Some BCI communication systems take the form of letter and word selection paradigms; common methods include letter selection by electroencephalography (EEG) using slow cortical potentials (SCP) [6–10], the P300 event-related potential (ERP) [11–17], steady state visual evoked potentials (SSVEP) [18–20], sensorimotor rhythms (SMR) [21–23] and event-related (de) synchronization (ERD/ERS) [24–32]. In addition, a recently developed spelling device was developed using electrocorticography (ECoG) for letter selection by ERD/ERS [33]. It is important to note that none of these systems actually predict intended speech, although typed statements may be voiced aloud using text-to-speech synthesis. There are two major disadvantages of this type of indirect communication system. First, while often very accurate, the letter selection rate can be as slow as one word/min, limiting a user's ability to converse fluently in real-time. Second, these systems are 'generic' in that they can be used for any item-selection task; thus, they ignore potentially valuable neurological information as well as speech-related constraints that may improve communication abilities.

Recent studies have tried to address these two problems and make BCI speech production more natural and fluent. In some studies, the neural activity related to speech imagery has been characterized using EEG and magnetoencephalography (MEG) [34–36]. In another example [37], a BCI was developed to perform direct word or phoneme prediction using EEG and speech motor imagery. This method attempts to address both issues mentioned previously; it is a direct classification of intended speech sound by neurological signal. However, the method used may not have the resolution needed to accurately represent all information needed for fluent speech production and has yet to be implemented for real-time use. A similar method is in development using ECoG to overcome the limitations of EEG resolution [38,39]. Other recent work has demonstrated it is possible to control a real-time speech synthesizer using intracortical microelectrodes implanted in speech areas of the motor cortex [40–43]. This method is analogous to continuously varying cursor control BCI applications, including systems using EEG [44–46], ECoG [47–49], and primate [50–56] and human [57–60] intracortical recordings. In addition, this approach synthesizes speech output in real-time as opposed to discrete approaches using slower-than-real-time text-to-speech synthesis. The real-time synthesis, EEG and ECoG methods described here are the first steps towards a true speech neural prosthesis, in contrast with item-selection BCIs applied to communication. The aim of such speech prostheses is to completely replace the vocal mechanism for individuals who are not able to properly use their existing biological vocal apparatus.

This article reviews the field of brain–computer interfacing, with particular focus on speech communication. The BCIs discussed in the following sections all seek to produce some extrinsic behavior such as word/letter selection or speech sound generation. These methods differ according to type of communication: indirect (e.g., spelling) versus direct (speech prosthesis), and mode of production: synchronous versus asynchronous. Recording methodology also differs (noninvasive vs invasive), as do decoding paradigm (discrete vs continuous) and feedback modality. We first clarify these terms before discussing

representative examples of current BCI applications. Table 1 summarizes the approaches described in the following sections.

BCI design principles

Recording technique

In BCI applications, both EEG and MEG have been used for noninvasive measurements of neurological activity, although EEG is more prevalent. The EEG signal describes neuroelectrical activity associated with currents flowing perpendicularly to the scalp surface, while MEG measures magnetic fields resulting from tangential currents. Both arise from the synchronous activity of millions of pyramidal cells in the cerebral cortex. Therefore, EEG and MEG are ill suited to describe the effects on the scale of an individual neuron but are well suited to describing the overall dynamics of neurons in the brain regions directly beneath recording sites.

Invasive recording techniques include intracranial recording (i.e., ECoG) and intracortical recording (i.e., multiunit extracellular microelectrodes). Both methods require neurosurgical operations to implant the recording device. ECoG is used to record synchronized activity neurons from electrode arrays on the cortical surface but avoids skull and scalp conductance, resulting in much higher signal-to-noise ratio (SNR) than EEG. Microelectrode recordings use electrodes implanted into the cortical surface, and differ from both EEG and ECoG. The signal represents a multiunit extracellular electrical potential resulting from the summed activity of many neurons near the electrode's recording tip. Microelectrodes record two main types of signal: single-unit waveforms (e.g., individual action potentials) and the local field potential (LFP), which represents the summed activity of many neurons near the electrode tip. Single units can be resolved from the multiunit recording by first sampling at a sufficient rate (>20 kHz), then filtering (e.g., 300–6000 Hz, bandpass), followed by spike detection and classification (see [61] for a review). Alternatively, the multiunit signal, if recorded with a low impedance electrode, can be lowpass-filtered (e.g., with 300 Hz cutoff) to obtain the LFP. Several multiunit extracellular electrode types exist for general neurophysiological research (e.g., [62–64]), although only two designs are available for use in chronic human implant applications, which is a requirement for speech BCIs. They include the Neurotrophic Electrode [65,66] and the Utah multielectrode array [67–69].

Production mode

Communicative BCI systems have two major production modes: synchronous and asynchronous. Synchronous BCIs rely on fixed-trial pacing, primarily to evoke neuroelectrical potentials in response to the external stimulus, while asynchronous BCIs rely solely on self-generated neurological activity and continuous-trial pacing. Thus, BCI operation proceeds at the user's own pace, typically using self-generated (compared with ERPs) responses. Therefore, a user with vision or hearing impairments can still communicate via asynchronous BCI but have difficulty with a synchronous device. Natural speech production is an asynchronous activity; no stimulus is required for neurologically normal populations to plan and execute speech movements. Therefore, a speech prosthesis may benefit from asynchronous designs.

Communication type

We will use the term indirect speech communication to refer to communicative BCIs that utilize neural activity unrelated to the act of speech production as the primary BCI control feature. This includes all techniques using visual evoked potentials (e.g., P300 and SSVEP), SCP and nonfacial sensorimotor ERD/ERS. In all cases, an intermediate step is required to translate the neural activity representing nonspeech information into a speech domain. It is important to

note that hand-related sensorimotor, ERD/ERS can be considered a direct method for typing-based BCIs rather than for speech production. Since this article is primarily concerned with BCIs for speech communication, even these hand sensorimotor methods will be considered indirect. By contrast, direct speech communication refers to BCI methods that utilize neural activity related innately to the act of speech production. These methods use the neurological activity present during attempted or imagined speech production as the basis for any BCI speech prediction. For direct methods, no intermediate mapping is required between cognitive states and speech output, which may dramatically increase word-production rates toward real-time fluent speech production. Second, direct prediction of attempted speech allows researchers to exploit known neural mechanisms for speech production and other relevant speech characteristics, in decoding and classification algorithms for speech prosthesis.

Target subject population

Brain-computer interfaces developed based on the design topics in the preceding paragraphs are ultimately subject to the needs of the target users who often vary in movement ability, perceptual ability and general neurological status. In general, noninvasive methods offer the widest range of potential users as surgery for invasive procedures is not required; not all users are good candidates for intracranial or intracortical electrode implants. One major consideration for invasive implantation is the presence of specific neurodegenerative diseases (e.g., ALS) that may undermine possible advantages of invasive measurement. Unfortunately, the distinct electrical signals of millions of neurons are lost when using noninvasive techniques.

Example BCI applications

The primary approach taken by most BCI applications is to use EEG for indirect speech communication via visual feedback, with varying production modes. This BCI model has broad availability to all target populations. This section reviews many BCI applications; some follow this model while others deviate. Table 1 summarizes all methods discussed with respect to the design principles in the previous section. Of particular interest are those BCI applications that perform asynchronous, direct speech prediction. These methods provide the greatest opportunity for true, fluent speech prosthesis. Some BCI users prefer to use slower communication speeds than the maximum possible speeds, presumably owing to the increased mental effort required to operate such devices at the fastest rate; also, users may sacrifice speed if a slower interface offers higher accuracy. However, fluent speech prostheses eliminate much of this effort by making BCI-based speech production an intuitive task in which subjects simply attempt to speak naturally.

EEG techniques

Thought translation device—One of the earliest attempts to restore communication by BCI, the Thought Translation Device (TTD), was developed by Birbaumer and colleagues [6,7]. Their BCI used biofeedback of the SCP to allow users to navigate a binary-tree spelling device. The SCP is an EEG amplitude modulation of oscillations less than 1 Hz, obtained primarily from the vertex (Cz) electrode location. Participants first learned to voluntarily modulate their SCP, either positive or negative amplitudes relative to baseline, through an intensive visual or audio feedback paradigm [6]. In the training paradigm, users watch a ball moving vertically in proportion to the user's SCP amplitude relative to baseline (e.g., down for positivity and up for negativity) [8,10]. Typically, users train for weeks or months before they can willfully modulate SCP amplitude, although one study [10] demonstrated ten out of 13 healthy subjects were able to produce statistically significant SCP amplitude modulations in a single session.

Following biofeedback training, users operate a synchronous binary-tree spelling program, participating in two phases of closed-loop spelling. First, users are instructed to directly copy presented words and letters by navigating the binary tree [8]. For instance, a positive SCP may yield selection of the 'left' sub-tree while a negative SCP yields 'right' sub-tree selections. Each sub-tree continues to branch off until a leaf node is chosen, representing the desired selection or letter. Promotion to the second phase of spelling, free spelling, required 75% accuracy in the copy spelling task, typically after many hundreds of training sessions [6,7, 10]. Spelling rates with the TTD are limited because the trial length needed to observe SCP changes is relatively long (two 2–4 s epochs; baseline and active phases, respectively [6–8, 10]) resulting in letter production rates near 0.5 letters/min [6]. Efforts have also been undertaken to improve spelling rates through increasing the degrees of freedom beyond binary choice by utilizing standard SCP as well as 'bipolar' SCP (bipolar recording from electrodes C3 and C4) [10].

The potential spelling accuracy and robust signal feature acquisition (i.e., SCP control after biofeedback training) afforded by the TTD have led to general success as a spelling tool to restore communication. Furthermore, both nondisabled and paralyzed populations have used this method with similar degrees of success. Unfortunately, it has a number of drawbacks, including intensive training time and very slow spelling rates. These rates are not likely to be improved upon as the nature of the SCP signal itself requires at least one second (owing to <1 Hz frequency component) per selection to accurately verify SCP amplitude modulations.

Sensorimotor rhythm: cursor selection—Another technique, developed by Wolpaw and colleagues, uses modulations in the SMR μ - and β -bands to control 1- and 2D cursor movement on a computer screen [21–23,44,45]. The μ -(8–12 Hz) and β -(18–25 Hz) SMR are EEG rhythms related to execution or imagery of motor movements. An extension of this BCI allows users to select communication-relevant items with the BCI cursor. Initial studies addressed the feasibility of synchronous 1D SMR control to answer yes/no questions in which users moved the cursor to the top of the screen to select a 'yes' answer and the bottom for 'no' [21]. Four subjects (one with ALS) were able to correctly answer 333–401 questions (over many sessions) with 78–93% accuracy, indicating that communication was possible using this type of BCI. A later study investigated the usefulness of a 1D cursor BCI in a typing protocol [22,23]. The spelling program consisted of four targets along the right side of a computer screen. Three targets contained letters and the fourth contained a backspace option. At the beginning of each spelling epoch, a cursor moved steadily across the screen from left to right while subjects controlled the vertical position; the final position of the cursor indicated the selected target. The authors report that users were able to spell up to one word/min (or five letters/min) [22].

P300 speller—The P300 is an ERP elicited when subjects observe a relatively rare target item among many other nontarget items. The signal itself is characterized by increases in the time-series amplitude, relative to rest or observations of nontarget items, approximately 300 ms after target stimulus observation. This increase in EEG positivity is most noticeable in parietal and occipital electrodes, though it can be generally observed at nearly any location over the scalp. The P300 was suggested initially as a feature for operation of a letter-spelling BCI by Farwell and Donchin [11] but has been more recently investigated by other research groups [12–17].

To elicit the P300 for use in a letter-spelling BCI paradigm, letters of the alphabet arranged in a matrix are displayed on a computer screen. Groups of letters (rows and columns), or randomly presented letters, are then individually highlighted in turn while the BCI user attends to the preselected target letter. According to the properties of the P300 ERP, a positive increase in scalp potentials should be elicited when the target letter is highlighted, since this is an infrequent

occurrence compared with other letter presentations [11]. The presentation sequence is repeated many times per selection for trial-averaging necessary to improve the P300 SNR for reliable detection. A recent study found that while single-letter and group highlighting both elicited expected P300 amplitudes relative to nontarget stimuli, the response was greatest when a single letter was highlighted as opposed to a row, column or random group [12].

A common form of the P300-based spelling BCI uses a 6×6 matrix that contains the 26 letters of the alphabet and ten additional items (often the numbers 0–9). Each row and column is illuminated in a cue-based, synchronous presentation paradigm for 100–175 ms (interstimulus interval [ISI]), totaling 12 intensifications, two containing the target item and ten containing non-target items. Variations in matrix size and ISI have been shown to affect P300 amplitude, accuracy and information rate [16,70]. Early implementations used midline electrodes (i.e., frontal [Fz], central [Cz] and parietal [Pz]), which had been shown to record robust P300 responses. Other studies reported that certain electrode locations are better correlated with target acquisition in P300 paradigms; significant classification improvements were reported when using a montage of electrodes that included the standard electrode set (Cz, Fz and Pz), as well as posterior electrodes PO7, PO8 and Oz [15]. The typical P300 spelling paradigm first involves copy spelling: users select letters by attending to items in the speller matrix while the sequence of intensifications is repeated. Following copy-spelling training, users are encouraged to spell letters and words of their choosing in a free-spelling task. Target letter classification is accomplished using stepwise linear discriminant analysis, which iteratively adds features for use in the linear discriminant function, which significantly contribute to the overall variance. For a more detailed description see [11].

Both nondisabled subjects [11–13,15,16] and disabled subjects (individuals with quadriplegia [12] and ALS [14,17]) have successfully learned to communicate with the P300 Speller with typically over 90% accuracy for healthy subjects and over 79% for disabled patients. Improvements in the P300 Speller algorithm have led to improved healthy-subject performance, from 2.3 characters/min [11] to 4.3 characters/min at 95% accuracy [12]. Quadriplegic and ALS patients have achieved rates of 3.2 characters/min at 95% accuracy [12] and 2.1 characters/min at 79% accuracy, respectively [14]. Importantly, the ALS patients were able to maintain this high level of performance for over 40 weeks, illustrating that the P300 Speller is viable for long-term use by ALS patient populations.

Steady state visual evoked potential—The SSVEP is a cortical oscillation elicited when users view a flickering stimulus. Specifically, spectral analysis of EEG in primarily visual areas (e.g., occipital electrodes: O1 and O2) reveals frequency components at integer multiples of the observed strobe frequency (base frequency and harmonics), with larger responses attributed to attended stimuli. Using the SSVEP, it is possible to create a BCI for spelling by presenting grids of items that flicker at distinct strobe frequencies [18]. With such a device, users attend to desired items, and EEG is collected according to a synchronous paradigm containing the elicited SSVEP for the target strobe frequency. Detection algorithms can be used to discriminate the target (i.e., largest) SSVEP response from nontarget items.

A pioneering study by Sutter proposed using the SSVEP to provide a keyboard selection interface using EEG [18]. In this implementation, 64 keys were available in an 8×8 grid in which each key flickered between red and green colors at different frequencies. Over 70 subjects with no neurological deficits evaluated the prototype and achieved typing speeds between 1 and 3 s/item [18]. Approximately 20 severely disabled persons with cerebral palsy or ALS also tested the system but were far less successful owing to electromyography contamination from the presence of uncontrolled neck movements. As a remedy, a single ALS patient was implanted with intracranial electrodes (i.e., ECoG) and used the resulting SSVEP system to communicate between 10 and 12 words/min (~ 1.2 s/character).

The following studies investigated the feasibility of SSVEP-based BCI for spelling applications, with particular interest in accuracy and information bandwidth using EEG [19, 20]. In one study participants used an SSVEP-based BCI for ‘spelling’ phone numbers by selecting among 13 buttons flickering at different strobe rates between 6 and 14 Hz. Eight out of 11 subjects correctly typed an 11-digit phone number at durations ranging from 45 to 135 s. The information transfer rate for a second task of 24-digit random number selection was between 19.22 and 55.69 bits/min (75–100% accuracy). In another study subjects spelled words using two control paradigms [20]. In the first paradigm, users selected rows and columns corresponding to the target in a 5×5 matrix by attending to one out of five flickering light emitting diodes (LEDs) with strobe frequencies between 13–17 Hz. Just two selections were required for letter selection using the grid layout. The second paradigm used a rhombus layout in which four of the LEDs represented UP, DOWN, LEFT and RIGHT selections relative to a center while the fifth LED indicated selection of the current letter. Using this layout, at least two selections were required for each letter. Nine out of 11 subjects completed both paradigms, spelling a three-word phrase (22 letters) with minimal errors (seven errors for the matrix paradigm and five errors for the rhombus paradigm). The mean information transfer rate was 28.4 and 30.6 bits/min for the matrix and rhombus spelling layouts, respectively.

Graz BCI—Pfurtscheller and colleagues at the Graz University of Technology developed a BCI for two-state classification using a mental imagery strategy [24,25]. The primary motivation for the Graz-BCI was to reduce training time needed to attain adequate two-state classification of EEG signals. Their solution used well known ERS/ERD of the SMR (namely μ - and β -rhythms). ERD are elicited during the planning and execution of motor movements while ERS are found at rest, or generally following ERD. During training, subjects are instructed to imagine specific motor movements (e.g., right vs left hand movement or both feet vs right or left hand movement) in response to a visual cue. The spectral power is computed in both the μ - and β -bands and then input to discrimination and classification algorithms for detection of class differences. The Graz-BCI is very similar to other SMR approaches used for prediction of two-class discrimination and computer cursor movements [21,44], and it has more recently been used to provide neural control over a binary-decision spelling device for nondisabled [26,27] and disabled populations [28].

Spelling with a ‘Virtual Keyboard’ (VK) is a recent addition to the Graz-BCI, using a two-class discrimination task – left versus right motor imagery – to select increasingly fine groups of letters from a binary letter selection algorithm [26]. Initially, 32 letters are displayed, half on the left side of a computer screen and half on the right, and users are instructed to imagine a motor movement, synchronized to the letter presentation, associated with the onscreen location (left or right side) of a desired letter. This is repeated until only one letter remains; then verification and error-correction selections are made. A total of six selections are required for a correct letter selection while at least 13 selections are needed to correct for erroneous selections [26]. In initial studies, subjects performed two copy-spelling tasks (totaling 44–46 letters). After the second VK spelling session, the subjects were able to select letters at rates ranging from 0.67 to 1.02 letters/min with more than 97% accuracy.

Follow-up studies specifically investigated the effects of increased numbers of discriminant classes [27], asynchronous letter selection [27], and disabled population performance [28]. Two of three nondisabled subjects, previously trained on a two-class VK, successfully used the three-class VK for copy-spelling. Their spelling rates improved (mean: 1.99 letters/min at 71% correct) compared with the two-class version. In a 22-week (2 days/week totaling 178 sessions) case study, a participant with cerebral palsy performed letter-spelling tasks with the two-class VK Graz-BCI [28]. The participant underwent standard Graz-BCI training, and used linear discriminant analysis for classification with incremental letter spelling (two-choice, single letter selection) followed by copy-spelling tasks. The participant demonstrated a learning

effect during the letter-spelling training stage with approximately 62% accuracy during the first ten sessions and 69% accuracy during the last ten sessions. In addition, the participant spelled 99 words of 4–8 letters each at an average spelling rate near one letter/min.

Berlin BCI—The Berlin-BCI is an alternative classification system for predicting intended binary state from EEG according to a machine learning perspective [29–32,71,72]. Like the Graz-BCI, the Berlin-BCI utilizes neurological activity related to executed or imagined movements. However, it uses a specialized spatial filtering technique (Common Spatial Pattern [CSP]) to optimize signal acquisition for maximum two-class discrimination [30,73]. Conversely, other methods rely on characteristic EEG responses (e.g., P300, SSVEP or ERD/ERS) and use simple spatial filters. All paradigms begin with a short CSP calibration period, in which subjects are instructed to attempt or imagine specific movements (e.g., moving left/right hand or foot). Next, typically a small number of the most discriminable CSPs are chosen and used for asynchronous binary state classification in the Berlin BCI.

This CSP-based selection technique was applied to spelling tasks using the Hex-o-Spell system. The device displays six hexagrams, containing 26 letters and four extra punctuation marks. The user controls an arrow inside the central region through CSP discrimination of motor imagery (e.g., right hand vs right foot movement). The arrow rotates clockwise in response to one imagined movement and grows in length for the other, eventually selecting a hexagram containing the desired symbol. This procedure is repeated until a single letter or symbol is selected (for further detail see [32]). In one study of spelling performance, two subjects, in a real-world environment, typed error-free (i.e., needed to erase or respell) words at rates between 2.3 and 7.6 letters/min [32].

Direct vowel classification: CSP—DaSalla and colleagues have extended an indirect CSP-based interface to attempt direct two-class vowel discrimination [37]. Subjects were instructed to perform vowel-speech imagery for CSP calibration and general BCI control; specifically, to imagine lip-rounding and mouth-opening (with vocalization for both) when presented with visual stimuli, in addition to a control condition involving no imagined movements. These articulatory configurations correspond well to the vowel sounds AA (hot) and UW (hoot) in three able-bodied subjects. CSPs were chosen to maximize discrimination between the three experimental conditions: AA versus rest, UW versus rest and AA versus UW. Though the recorded EEG is related directly to the motor imagery of vowel production tasks, the CSPs are not guaranteed to yield speech-related information. For instance, in pairwise comparisons between the two vowel conditions and the rest condition, the spatial patterns show bilateral activity recorded by sensorimotor electrodes, but when vowel conditions are compared directly, only the spatial pattern for AA demonstrated sensorimotor activity. This artifact is a property of the algorithm and is very useful for discrimination of maximally different spatial filters, but obscures interpretation of the differences in EEG between vowel conditions. Offline classification analysis indicated that two out of three subjects' spatial patterns were discriminated above chance for the vowel comparison AA versus UW, although all were above chance for vowel versus rest conditions (UW vs rest being highest). Overall, this method for vowel classification resulted in classification accuracies between 68 and 78%.

Intracranial techniques

Sensorimotor ERD/ERS: ECoG—Electrocorticography is an increasingly popular method for obtaining high SNR electrophysiology from nonparalyzed subjects, typically with severe epilepsy. A number of studies have begun to investigate using ECoG for BCI applications involving 1- and 2-D cursor control [47–49], but few have studied the effectiveness of ECoG for spelling-based BCI communication devices [33]. In a study of five patients with acute ECoG preparations for the identification of epileptic foci prior to surgery, Hinterberger and colleagues

investigated ECoG control of an ERD/ERS-style BCI for binary letter selection [33]. All patients had electrode arrays partially placed over the primary motor and premotor cortex. Initial training was similar to that of the Graz-BCI paradigm in which subjects first performed predefined motor imagery (finger vs tongue movements) cued by a visual stimulus. For this imagery, spatially distinct ERD should manifest in electrode recordings over finger/hand areas for the finger movement and over tongue areas for tongue movement. Three out of the five patients successfully performed between 157 and 244 trials of copy-spelling in a single session (one patient participated in two sessions) with accuracies between 64 and 88% correct (mean: 76%) and spelling rates between 0.32 and 0.82 letters/min (mean: 0.41 letters/min).

Direct word/phoneme classification: ECoG—The existing ECoG BCI literature and direct speech classification studies has led to preliminary investigations of ECoG related to both actual and imagined speech productions [38,39]. In these studies, patients with severe epilepsy were temporarily implanted with ECoG arrays and performed multiple motor tasks including word and phoneme production. In one study, nine subjects' ECoG during word production was classified offline according to four vowel and four consonant groups [38]. In both studies, the authors conclude that the ECoG signal contains sufficient information to reliably discriminate between certain speech sounds. This work is in its initial stages, and as such has led to limited results. However, this is a promising and potentially successful approach to discrete, direct speech prediction.

Motor cortex extracellular microelectrode: cursor selection—Extracellular microelectrode recordings have been long used to investigate motor-cortical activity in primates in response to motor execution [74], and more recently for operation of BCIs [50–56,75]. Kennedy and colleagues performed the first human chronic microelectrode implant study using the Neurotrophic Electrode implant (see [66,77] for further details) involving two-class discrimination by microelectrode recordings from the motor cortex of an individual with ALS [76]. Subsequent studies investigated the feasibility of a motor cortical BCI for cursor control [78,79]. In these studies the LFP was isolated and input into an algorithm that produced a 1D cursor movement. Over many training sessions, the implant recipient learned to spell by moving a cursor over a virtual keyboard, achieving information transfer rates of approximately three letters/min (accuracy not reported). A motor-cortical BCI for cursor-controlled item selection has been attempted using the Utah microelectrode array [57,59,60]. Preliminary results have shown implant recipients have achieved 73–95% accuracy with mean acquisition time of 2.5 s [60]. Although this system was not used directly for spelling, it was used for general environment interaction and can be adapted for spelling purposes.

Motor cortex extracellular microelectrode: direct phoneme prediction—Kennedy and colleagues made initial investigations into phoneme prediction using intracortical microelectrode recordings, [80–82] with the goal of predicting 39 English phonemes (and subsets) from single and multiunit activity using the Neurotrophic Electrode implant [65,66]. In these studies, a single subject with LIS due to brainstem stroke, participated in various attempted speech production (i.e., imagery) tasks. He was implanted in the left speech motor cortex (specifically left precentral gyrus, on the primary motor cortex and premotor cortex border) after a preoperative fMRI study, involving attempted picture naming and word repetition, indicated this region produced the greatest neural response. In these studies the subject repeated phonemes after auditory stimulus presentation but received no feedback regarding the BCI prediction. In two of the studies, advanced pattern recognition algorithms (i.e., support vector machines) attained high levels of accuracy through discrete classification of selected phonemes [80,81]. The third study attempted to predict the formant frequencies corresponding to three vowel sounds (AA: hot, IY: heat and UW: hoot) and used linear discriminant analysis for classification into vowel categories. Briefly, formant frequencies are

the resonant frequencies of the vocal tract, which are related to the overall volume of the oral cavity and modified by movements of the speech articulators. Formant frequency prediction can also be used for direct speech synthesis of vowel sounds in real-time, providing instantaneous auditory feedback to the BCI user. Unfortunately, formant analysis and prediction cannot be easily used to produce consonant sounds, which are defined by places and manners of articulatory closures. For these sounds – and complete fluent speech – prediction of articulatory configurations are needed in combination with continuous articulator-based speech synthesizers (e.g., [83]).

A recent study extended previous studies and investigated the properties of speech motor cortex during attempted speech productions, with respect to formant frequencies, and the feasibility of real-time speech prosthesis via formant frequency prediction and artificial speech synthesis [40,41]. In this study, the same participant listened to vowel sequences in a center-out paradigm (see [74] for the original center-out paradigm) in which a neutral vowel (AH: hut) was the center and three corner vowels (AA, IY and UW) were peripherally located. An initial phase of cued-stimulus speech motor imagery was used for decoding filter calibration. In particular, the firing rate of recorded single- and multi-units were computed, and a Kalman filter neural decoder was constructed to predict the 2D stimulus formant frequency trajectory being mimicked [57,84]. Following model calibration, the subject performed a listen-and-repeat protocol of center-out vowel-vowel sequences (e.g., AH-AA, AH-UW and AH-IY). The neural prosthesis transformed firing rates into formant frequencies for instantaneous vowel synthesis and computer playback. Over many sessions, the participant correctly produced the specified stimulus between 45% (early trials/session) and 70% (late trials/session). According to a standard definition of information transfer rate (i.e., bit rate), the within-session performance of the speech prosthesis was computed between 0.57 and 6.97 bits/min (computed based on reported accuracy, targets and movement time), from early to late trials. As mentioned earlier in Table 1, this information rate is approximately equal to the spelling rate as only a single ‘selection’ is needed.

Discussion

The BCI methods described in this article all provide a means for individuals with profound paralysis to communicate. Each has advantages and constraints, and may be more or less useful depending on a user’s unique conditions.

Recording methodology

Noninvasive approaches based on EEG have a number of significant benefits compared with invasive techniques. Most importantly, they do not require surgery to implant neurophysiological recording electrodes. This enables testing and development with able-bodied, neurologically normal subjects prior to use with intended disabled subjects. In addition, EEG and MEG record neural activity from the entire cortical surface simultaneously, enabling generalized methods for extracting the maximum information for use in BCI control (e.g., CSP-method). Furthermore, these devices have an established history of successfully restoring communication for disabled users through spelling applications.

By contrast, invasive methods for chronic human use are relatively recent additions to the BCI literature and are, as yet, unproven over large user populations. However, they have shown great promise for simultaneous and accurate classification and decoding of many ($n > 2$) speech sounds. In addition, M/EEG recordings are based on combined neural sources of activity, resulting in poorer spatial resolution and limiting the number of degrees of freedom that can be used for decoding and classification systems. It is possible that BCIs for fluent speech production require access to a larger number of independent neural units (e.g., singleunit, multiunits or ECoG).

Another invasive implant consideration is the implantation site, which will highly interact with the decoding methodology. For instance, a speech motor cortical implant is designed to intercept and interpret the final motor commands issued to the vocal articulators. A continuous filter decoder may be most appropriate for such fluid behavior. However, other potential implant locations, such as the supplementary motor area and inferior frontal gyrus (i.e., Broca's area) may encode more discrete representations of speech sounds. Implants in these regions may be better served by discrete classification methods for phoneme prediction. Last, another major factor influencing implantation site is a physical constraint; specifically the ability to surgically access the intended brain region.

Production mode

Communication interfaces using stimulus-synchronized neural activity acquisition (e.g., P300, SSVEP and TTD), provide some of the highest accuracy information transfer rates with proven usability by disabled populations. However, all stimulus-synchronized methods require that users have high degrees of sensory perception (e.g., vision and hearing), which is often impaired in people with LIS. Consequently, some of the most recent advances in BCI communication technology have derived from asynchronous neural signal acquisition, using self-generated brain activity for control of external devices. Although users still need some perceptual abilities in order to monitor communication monitoring errors, asynchronous production methods eliminate the need for slow feedback loops and permit fully feedforward-based communication. Unfortunately, some BCI users (e.g., those suffering from ALS) may have difficulty using asynchronous systems because the most common self-generated rhythms involve motor imagery of arm, hand or speech movements. For these users, a comprehensive, synchronous feedback-loop BCI may be the only means for reliable communication.

Communication type

Despite overwhelming success at restoring communication to disabled individuals, indirect communication devices have a number of disadvantages. While classification accuracy is often high, the information rate (or characters/min) is unsatisfactorily slow for any attempt at fluent speech-sound production or real-time communication. For instance, the TTD interface permits only one letter selection every 2 min, meaning it would take a full hour to spell out 30 words or two to three sentences, while other methods commonly report less than six letters/min. Anecdotal reports suggest that chronic, noninvasive BCI users will choose slower spelling devices if the mental effort needed to control such devices causes the user to lose accuracy at faster speeds. Direct decoding methods eliminate the need to select single letters, phonemes or words; instead users simply think about the word/sound they want to produce. In addition, direct speech predictions can be potentially produced instantaneously with users' attempted or imagined speech, thereby allowing production rates similar to natural speech [41]. With such devices, BCI users would not have to choose between speed and accuracy, or effort.

Another consideration for direct speech-sound production devices is the method of prediction: discrete versus continuous. Most direct methods rely on classification of neural activity into a discrete word, syllable or phoneme group (i.e., vowels or consonants). Unfortunately, discrete approaches suffer one major drawback; the size of the dictionary needed to store all words, syllables and phonemes may become prohibitively or intractably large. For instance, there are few phonemes (in English), including vowels and most consonant-vowel pairs. However, there are many more syllables (e.g., all consonant-vowel-consonant triples), numbering into the hundreds. Beyond the level of syllables, there are thousands of words, which are required for fluent speech vocabularies. Any discrete speech prosthesis performing classifications at levels higher than the phoneme must account for this combinatorial consideration. Conversely, continuous filtering methods for neural decoding can reduce the degrees of freedom by choosing the output modality in a sufficiently small space. For instance, the device in uses a

low-dimensional (2D) auditory space, related to the movements of the vocal articulators to provide access to all steady, monophthongal vowels in English (although production of just three vowels was tested) [41,42]. More complex vowel sounds can be defined by movements within the continuous space (e.g., diphthongs). For instance, the vowel AY (high) is represented by a formant trajectory between the vowel sounds AA and IY in a 2D formant-frequency space; therefore, correct classification depends on producing the trajectory that defines the vowel sound. This dramatically increases the number of speech sounds that can be produced with continuous speech BCI and increases the potential information rate without changing the required number of continuous degrees of freedom. Similarly, one can design a BCI to control a low-dimensional representation of the vocal articulators for use in an articulatory speech synthesizer (e.g., [83]). Furthermore, an articulator-based BCI is required for full vowel and consonant production as acoustic speech BCIs are not capable of production consonants. Continuous decoding BCIs, compared with discrete classification, must make a tradeoff between the number of dictionary items (e.g., phonemes) and accurate control of continuous variables.

Target population

Many of the BCIs discussed in this article were designed specifically for disabled populations. In general, persons with non-neurodegenerative LIS (e.g., brain–stem stroke, cerebral palsy and spinal injury) should be capable of operating all BCIs discussed. Some limitations arise if a user has confounding complications, such as perceptual deficits (e.g., visual or auditory impairments). Such complications may eliminate certain BCI techniques that for instance, rely primarily on visual information, such as the P300 Speller and SSVEP device. Users with such perceptual deficits may benefit most from asynchronous spelling or communication, which do not require a strict perception–action protocol. However, most asynchronous methods rely on endogenous neural activity related to voluntary behavior, often related to imagined motor execution, which may be degraded in users with motor neuron diseases (e.g., ALS). Unfortunately, no method has yet been shown to be effective for patients with no remaining communication. However, it is possible, that advances in BCI technology may one day aid such patients.

Conclusion

Design of neural communication prostheses is an active area of investigation for many research groups. The most common form of communication device uses EEG measurements as a neurological signal to control spelling devices. Many variants of EEG-based spelling devices exist, including those using the P300 ERP and SSVEP synchronous feedback responses for selections of letters from a visual display. Other variants use willful modulation of the SCP and motor imagery for selection of letters from binary spelling devices. Cumulatively, these EEG studies indicate that it is practical to use noninvasive neurophysiological methods to control spelling devices. One major drawback of EEG-based, noninvasive methods is slow spelling rates, often in the order of a few letters/min. Dramatically faster spelling rates are required for fluent verbal communication. Slow spelling rates do allow individuals to communicate thoughts, desires and questions, but they cannot participate in natural social interactions, particularly those involving multiple conversants. Importantly, although slow interfaces are infinitely preferable to none, the social handicap of slow speech production may cause disabled users to withdraw from social interactions in frustration. Investigators have attempted to improve upon this deficit through usage of intracranial surface electrode grids (ECoG) that provide increased SNR, resulting in augmented discrimination of neural activity. These methods have largely been limited to motor imagery strategies for selection of letters through binary tree traversal, with impressive early successes.

Recent studies have been focused on performing direct speech sound prediction from various neurological signals. Researchers have used EEG for offline discrimination between two vowel sounds using a spatial filtering method [37]. Others have begun similar investigations of speech sound classification using ECoG [38,39]. Another study investigated using speech motor cortical activity in a continuous speech sound production device with instantaneous auditory feedback [41,42]. This study confirmed the presence of speech-related auditory information in the activity of recorded neural units in the speech motor cortex, and provided proof-of-concept for real-time, continuous speech prostheses. All direct methods for speech-sound prediction, as well as invasive techniques for BCI are in early stages of investigation, but show great promise for improved access to speech communication devices for severely paralyzed persons.

Expert commentary

The field of brain–computer interfacing has expanded dramatically over the past few decades. The primary purpose of BCI technology is to restore communication to severely paralyzed humans through a neural interface. A primary distinction between BCI technologies exists between recording methods (invasive vs noninvasive), production mode (synchronous vs asynchronous), and decoding method (discrete vs continuous). A further distinction made in this article is between communication type (direct vs indirect).

The most common BCI applications use noninvasive methods for recording neural activity and indirect means of communication. That is, some other behavior (e.g., nonspeech motor activity) or evoked response (e.g., P300 or visual evoked potentials) are the primary means for control of a communication device. Recent studies have begun investigating the feasibility of performing direct speech sound prediction and production via discrete classification and continuous filtering methods. These methods may provide a more intuitive interface for BCI control, alleviating some effort required with other common methods and hopefully move toward fluent speech prostheses.

Future studies are poised to expand the initial results provided in this article using more detailed speech articulatory information for continuous and discrete prediction, and real-time synthesis of consonants and vowels. In this way, multiple phonemes can be sequenced and generated through manipulation of a speech prosthesis that mimics and replaces the neuromuscular mechanisms for control of the vocal tract. It is possible that only intracranial techniques have the required specificity and resolution to capture fine enough neurological activity related to the activation of the vocal tract to effectively control such an artificial speech mechanism. In addition, EEG approaches will remain the only viable option for certain patient populations. Therefore, EEG-based BCI research will continue to improve current designs and reduce the training time needed to accurately control communication devices while intracranial BCI research proceeds toward a solution for direct speech sound production.

Five-year view

In 5-years time, new speech prosthesis investigations will report on the feasibility of decoding speech articulatory information and discrete phoneme prediction from chronic electrode implantation, over and within the human motor cortex. The goal of these studies will be to provide consonant and vowel production capabilities to people with severe paralysis, although may have limited utility for patients suffering from neurodegenerative diseases (e.g., ALS). Successful articulatory prediction will result in immediate capabilities of syllable level production by BCI recipients as they require only one vowel and one consonant; such two-phoneme synthesis has already been reported for vowel–vowel sequences.

Electroencephalography approaches for speech communication for spelling devices will not be discontinued in 5-years time. Rather, this line of research will be directed at obtaining optimal feature selection and classification methods rather than proving the feasibility and applicability of neurologically based control over spelling devices. Such investigations will probably result in improved information transfer rates and faster training times. Future EEG studies of speech-production-based interfaces for communication in able-bodied populations will derive from existing communication interfaces (both spelling and direct speech production) and may be used to prototype new methods for use in improved intracranial speech communication BCIs.

Key issues

- Brain–computer interfaces for speech communication have been achieved through primarily noninvasive means, although intracranial and intracortical attempts are increasing in frequency.
- Electroencephalography techniques employ two major strategies: synchronous perceptual feedback based control (P300 and steady state visual evoked potentials) versus asynchronous self-generated cortical signals (slow cortical potentials and event-related [de]synchronization of sensorimotor activity).
- Electroencephalography methods tend to be slow, permitting communication rates of, at most, a few letters/min; although this is acceptable to current users.
- Typically methods provide speech communication through selection of neural signals from indirect functional networks not related to speech.
- Recent research has been aimed at providing direct classification and production of speech sounds (e.g., phonemes).
- Microelectrode techniques that decode neural activity into acoustic correlates of intended speech production have been used for direct control over a speech synthesizer for production of vowel sounds in real-time.
- Initial observations of electrocorticography during speech production have demonstrated classification of individual phonemes is possible; supporting the notion that intracranial techniques will aid direct speech prediction.

Acknowledgments

The authors wish to thank Sara Rosenbaum for her assistance in the preparation of this manuscript.

This work was supported by NIH/NIDCD grants R01DC002852, R01DC007683 and by CELEST, an NSF Science of Learning Center (NSF SMA-0835976).

References

1. Plum F, Posner JB. The diagnosis of stupor and coma. *Contemp. Neurol. Ser* 1972;10:1–286. [PubMed: 4664014]
2. Haig AJ, Katz RT, Sahgal V. Mortality and complications of the locked-in syndrome. *Arch. Phys. Med. Rehabil* 1987;68:24–27. [PubMed: 2432852]
3. Katz RT, Haig AJ, Clark BB, DiPaola RJ. Long-term survival, prognosis, and life-care planning for 29 patients with chronic locked-in syndrome. *Arch. Phys. Med. Rehabil* 1992;73:403–408. [PubMed: 1580764]

4. Doble JE, Haig AJ, Anderson C, Katz R. Impairment, activity, participation, life satisfaction, and survival in persons with locked-in syndrome for over a decade: follow-up on a previously reported cohort. *J. Head Trauma Rehabil* 2003;18:435–444. [PubMed: 12973273]
5. Smith E, Delargy M. Locked-in syndrome. *BMJ* 2005;330:406–409. [PubMed: 15718541]
6. Birbaumer N, Ghanayim N, Hinterberger T, et al. A spelling device for the paralyzed. *Nature* 1999;398:297–298. [PubMed: 10192330]
7. Birbaumer N, Kubler A, Ghanayim N, et al. The thought translation device (TTD) for completely paralyzed patients. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2000;8:190–193.
8. Hinterberger T, Kübler A, Kaiser J, Neumann N, Birbaumer N. A brain-computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device. *Clin. Neurophysiol* 2003;114:416–425. [PubMed: 12705422]
9. Birbaumer N, Hinterberger T, Kübler A, Neumann N. The thought-translation device (TTD): neurobehavioral mechanisms and clinical outcome. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2003;11:120–123. [PubMed: 12899251]
10. Kübler A, Kotchoubey B, Hinterberger T, et al. The thought translation device: a neurophysiological approach to communication in total motor paralysis. *Exp. Brain Res* 1999;124:223–232. [PubMed: 9928845]
11. Farwell L, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Clin. Neurophysiol* 1988;70:510–523.
12. Donchin E, Spencer K, Wijesinghe R. The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2000;8:174–179.
13. Krusienski DJ, Sellers EW, Cabestaing F, et al. A comparison of classification techniques for the P300 Speller. *J. Neural Eng* 2006;3:299–305. [PubMed: 17124334]
14. Nijboer F, Sellers E, Mellinger J, et al. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clin. Neurophysiol* 2008;119:1909–1916. [PubMed: 18571984]
15. Krusienski DJ, Sellers EW, McFarland DJ, Vaughan TM, Wolpaw JR. Toward enhanced P300 speller performance. *J. Neurosci. Methods* 2008;167:15–21. [PubMed: 17822777]
16. Sellers EW, Krusienski DJ, McFarland DJ, Vaughan TM, Wolpaw JR. A P300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance. *Biol. Psychol* 2006;73:242–252. [PubMed: 16860920]
17. Sellers EW, Donchin E. A P300-based brain-computer interface: initial tests by patients ALS. *Clin. Neurophysiol* 2006;117:538–548. [PubMed: 16461003]
18. Sutter EE. The brain response interface: communication through visually-induced electrical brain responses. *J. Microcomputer. App* 1992;15:31–45.
19. Cheng M, Gao X, Gao S, Xu D. Design and implementation of a brain-computer interface with high transfer rates. *IEEE Trans. Biomed. Eng* 2002;49:1181–1186. [PubMed: 12374343]
20. Friman, O.; Luth, T.; Volosyak, I.; Graser, A. Spelling with steady-state visual evoked potentials; 3rd International IEEE/EMBS Conference on Neural Engineering, 2007. CNE 07; 2007. p. 354-357.
21. Miner LA, McFarland DJ, Wolpaw JR. Answering questions with an electroencephalogram-based brain-computer interface. *Arch. Phys. Med. Rehabil* 1998;79:1029–1033. [PubMed: 9749678]
22. Vaughan, TM.; McFarland, DJ.; Schalk, G.; Sarnacki, WA. EEG-based brain-computer interface: development of a speller; Neurocience Meeting Planner 2001; San Diego, CA, USA: 2001.
23. Vaughan T, McFarland D, Schalk G, et al. The wadsworth BCI research and development program: at home with BCI. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2006;14:229–233. [PubMed: 16792301]
24. Pfurtscheller G, Neuper C. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE* 2001;89:1123–1134.
25. Neuper, C.; Müller-Putz, GR.; Scherer, R.; Pfurtscheller, G. Event-Related Dynamics of Brain Oscillations. Elsevier; 2006. Motor imagery and EEG-based control of spelling devices and neuroprostheses; p. 393-409.
26. Obermaier B, Muller G, Pfurtscheller G. “Virtual keyboard” controlled by spontaneous activity EEG. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2003;11:422–426. [PubMed: 14960119]

27. Scherer R, Muller G, Neuper C, Graimann B, Pfurtscheller G. An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate. *IEEE Trans. Biomed. Eng* 2004;51:979–984. [PubMed: 15188868]
28. Neuper C, Müller GR, Kübler A, Birbaumer N, Pfurtscheller G. Clinical application of an EEG-based brain-computer interface: a case study in a patient with severe motor impairment. *Clin. Neurophysiol* 2003;114:399–409. [PubMed: 12705420]
29. Blankertz B, Dornhege G, Lemm S, Krauledat M, Curio G, Müller K. The Berlin brain–computer interface: machine learning based detection of user specific brain states. *J. Univers. Comput. Sci* 2006;12:581–607.
30. Blankertz B, Tangermann M, Popescu F, et al. The Berlin brain–computer interface. *Lect. Notes Comput. Sci* 2008;79:5050.
31. Blankertz B, Losch F, Krauledat M, Dornhege G, Curio G, Müller K. The Berlin brain–computer interface: accurate performance from first-session in BCI-naïve subjects. *IEEE Trans. Biomed. Eng* 2008;55:2452–2462. [PubMed: 18838371]
32. Blankertz, B.; Krauledat, M.; Dornhege, G.; Williamson, J.; Murray-Smith, R.; Müller, K. *Universal Access in Human-Computer Interaction. Ambient Interaction*. Berlin, Heidelberg: Springer-Verlag; 2007. A note on brain actuated spelling with the Berlin brain–computer interface; p. 759-768.
33. Hinterberger T, Widmann G, Lal TN, et al. Voluntary brain regulation and communication with ECoG-signals. *Epilepsy Behav* 2008;13:300–306. [PubMed: 18495541]
34. Suppes P, Lu Z, Han B. Brain wave recognition of words. *Proc. Natl Acad. Sci. USA* 1997;94:14965–14969. [PubMed: 9405722]
35. Suppes P, Han B. Brain-wave representation of words by superposition of a few sine waves. *Proc. Natl Acad. Sci. USA* 2000;97:8738–8743. [PubMed: 10890906]
36. Guimaraes M, Wong DK, Uy E, Grosenick L, Suppes P. Single-trial classification of recordings MEG. *IEEE Trans. Biomed. Eng* 2007;54:436–443. [PubMed: 17355055]
37. DaSalla CS, Kambara H, Sato M, Koike Y. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Netw* 2009;22:1334–1339. [PubMed: 19497710]
38. Schalk, G.; Barbour, D.; Leuthardt, EC.; Pei, X. Decoding spoken and imagined word groups using electrocorticographic signals in humans; *Neuroscience Meeting Planner 2009; IL, USA: 2009*.
39. Leuthardt, EC.; Freudenberg, Z.; Gaona, C., et al. Microscale electrocorticographic recording from human cortex and neuroprosthetic implications; *Neuroscience Meeting Planner 2009; IL, USA: 2009*.
40. Brumberg, JS.; Kennedy, PR.; Guenther, FH. Artificial Speech Synthesizer Control by Brain-Computer Interface; *Proceedings of the 10th Annual Conference of the International Speech Communication Association; Brighton, UK: International Speech Communication Association; 2009*.
41. Guenther FH, Brumberg JS, Wright EJ, et al. A wireless brain-machine interface for real-time speech synthesis. *PLoS ONE* 2009;4:E8218. [PubMed: 20011034]
42. Brumberg JS, Nieto-Castanon A, Kennedy PR, Guenther FH. Brain-computer interfaces for speech communication. *Speech Commun* 2010;52:367–379. [PubMed: 20204164]
43. Kennedy, PR. Comparing electrodes for use as cortical control signals: tiny tines, tiny wires or tiny cones on wires: Which is best?. Boca Raton: CRS/Taylor and Francis: *The Biomedical Engineering Handbook*; 2006.
44. Wolpaw JR, McFarland DJ. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl Acad. Sci. USA* 2004;101:17849–17854. [PubMed: 15585584]
45. Wolpaw J, McFarland D, Vaughan T. Brain-computer interface research at the Wadsworth Center. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2000;8:222–226.
46. Wolpaw J, McFarland D, Vaughan T, Schalk G. The Wadsworth Center brain-computer interface (BCI) research and development program. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2003;11:1–4. [PubMed: 12797719]
47. Leuthardt EC, Schalk G, Wolpaw JR, Ojemann JG, Moran DW. A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng* 2004;1:63–71. [PubMed: 15876624]
48. Schalk G, Kubánek J, Miller KJ, et al. Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. *J. Neural Eng* 2007;4:264–275. [PubMed: 17873429]

49. Schalk G, Miller KJ, Anderson NR, et al. Two-dimensional movement control using electrocorticographic signals in humans. *J. Neural Eng* 2008;5:75–84. [PubMed: 18310813]
50. Serruya MD, Hatsopoulos NG, Paninski L, Fellows MR, Donoghue JP. Instant neural control of a movement signal. *Nature* 2002;416:141–142. [PubMed: 11894084]
51. Paninski L, Fellows MR, Hatsopoulos NG, Donoghue JP. Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *J. Neurophysiol* 2004;91:515–532. [PubMed: 13679402]
52. Taylor DM, Tillery SI, Schwartz AB. Direct cortical control of 3D neuroprosthetic devices. *Science* 2002;296:1829–1832. [PubMed: 12052948]
53. Velliste M, Perel S, Spalding MC, Whitford AS, Schwartz AB. Cortical control of a prosthetic arm for self-feeding. *Nature* 2008;453:1098–1101. [PubMed: 18509337]
54. Chapin JK, Moxon KA, Markowitz RS, Nicolelis MAL. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nat. Neurosci* 1999;2:664–670. [PubMed: 10404201]
55. Wessberg J, Stambaugh CR, Kralik JD, et al. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* 2000;408:361–365. [PubMed: 11099043]
56. Carmena JM, Lebedev MA, Crist RE, et al. Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol* 2003;1:193–208.
57. Kim, S.; Simeral, JD.; Hochberg, LR.; Donoghue, JP.; Friehs, GM.; Black, MJ. Multi-state decoding of point-and-click control signals from motor cortical activity in a human with tetraplegia; 3rd International IEEE/EMBS Conference Neural Engineering 2007. CNE'07; 2007. p. 486-489.
58. 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;579:603–611. [PubMed: 17272345]
59. Hochberg, LR.; Simeral, JD.; Kim, S., et al. More than 2 years of intracortically-based cursor control via a neural interface system; Neuroscence Meeting Planner 2008; Washington, DC, USA: 2008.
60. Hochberg LR, Serruya MD, Friehs GM, et al. Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 2006;442:164–171. [PubMed: 16838014]
61. Lewicki MS. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation Neur. Syst* 1998;9:R53–R78.
62. Wise KD, Angell JB, Starr A. An integrated-circuit approach to extracellular microelectrodes. *IEEE Trans. Biomed. Eng* 1970;17(3):238–247. [PubMed: 5431636]
63. Hoogerwerf A, Wise KA. Three-dimensional microelectrode array for chronic neural recording. *IEEE Trans. Biomed. Eng* 1994;41:1136–1146. [PubMed: 7851915]
64. Nicolelis MAL, Dimitrov D, Carmena JM, et al. Chronic, multisite, multielectrode recordings in macaque monkeys. *PNAS* 2003;100:11041–11046. [PubMed: 12960378]
65. Kennedy PR. The cone electrode: a long-term electrode that records from neurites grown onto its recording surface. *J. Neurosci. Methods* 1989;29:181–193. [PubMed: 2796391]
66. Bartels JL, Andreasen D, Ehirim P, et al. Neurotrophic electrode: method of assembly and implantation into human motor speech cortex. *J. Neurosci. Methods* 2008;174:168–176. [PubMed: 18672003]
67. Jones K, Campbell P, Normann R. A glass/silicon composite intracortical electrode array. *Ann. Biomed. Eng* 1992;20:423–437. [PubMed: 1510294]
68. Maynard EM, Nordhausen CT, Normann RA. The Utah Intracortical electrode array: a recording structure for potential brain-computer interfaces. *Clin. Neurophysiol* 1997;102:228–239.
69. Rousche PJ, Normann RA. Chronic recording capability of the Utah Intracortical Electrode Array in cat sensory cortex. *J. Neurosci. Methods* 1998;82:1–15. [PubMed: 10223510]
70. Allison B, Pineda J. ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2003;11:110–113. [PubMed: 12899248]
71. Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller K. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Magazine Proc* 2008;25:581–607.
72. Krauledat M, Tangermann M, Blankertz B, Müller K. Towards zero training for brain-computer interfacing. *PLoS ONE* 2008;3:E2967. [PubMed: 18698427]

73. Ramoser H, Müller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2000;8:441–446.
74. Georgopoulos AP, Schwartz AB, Kettner RE. Neuronal population coding of movement direction. *Science* 1986;233:1416–1419. [PubMed: 3749885]
75. Wu W, Gao Y, Bienenstock E, Donoghue JP, Black MJ. Bayesian population decoding of motor cortical activity using a kalman filter. *Neural Comput* 2006;18:80–118. [PubMed: 16354382]
76. Kennedy PR, Bakay RAE. Restoration of neural output from a paralyzed patient by direct brain connection. *Neuroreport* 1998;9:1707–1711. [PubMed: 9665587]
77. Kennedy PR, Mirra SS, Bakay RAE. The cone electrode: ultrastructural studies following long-term recording in rat and monkey cortex. *Neuroscience Lett* 1992;142:89–94.
78. Kennedy PR, Bakay RAE, Moore MM, Adams K, Goldwithe J. Direct control of a computer from the human central nervous system. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2000;8:198–202.
79. Kennedy PR, Kirby T, Moore MM, King B, Mallory A. Computer control using human intracortical local field potentials. *IEEE Trans. Neural. Syst. Rehabil. Eng* 2004;12:339–344. [PubMed: 15473196]
80. Wright, EJ.; Andreasen, DS.; Bartels, JL., et al. Human speech cortex long-term recordings: neural net analyses; Neuroscience Meeting Planner 2007; CA, USA: 2007.
81. Miller, LE.; Andreasen, DS.; Bartels, JL., et al. Human speech cortex long-term recordings: Bayesian analyses; Neuroscience Meeting Planner 2007; CA, USA: 2007.
82. Brumberg, JS.; Andreasen, DS.; Bartels, JL., et al. Human speech cortex long-term recordings: formant frequency analyses; Neuroscience Meeting Planner 2007; CA, USA: 2007.
83. Maeda, S. *Speech Production and Speech Modeling*. Boston: Kluwer Academic Publishers, USA; 1990. Compensatory articulation during speech: evidence from the analysis and synthesis of vocal tract shapes using an articulatory model.
84. Kalman RE. A new approach to linear filtering and prediction problems. *J. Basic Biol. Eng* 1960;82:35–45.

Table 1

Summary of brain-computer interface methods for communication.

Method	Subject population	Neural recording	Communication modality	Performance	Mean production rates
Thought translation device	Healthy and disabled (ALS-LIS)	EEG	Letter spelling device	75–100%	0.5 letters/min
P300 speller	Healthy and disabled	EEG	Letter spelling device	62% (disabled), 80–90% (abled)	2.1–3.2 letters/min (disabled); 4.3 letters/min (able)
Steady state visual evoked potentials	Healthy	EEG	Letter/digit selection device	68–100%	19.22–55.69 bits/min
Graz-BCI (motor imagery)	Healthy and disabled	EEG	Two-class motor imagery	>71% (able), 70% (disabled)	1.99 letters/min (able), 1 letter/min (disabled)
Berlin-BCI (common spatial patterns)	Healthy	EEG	Two-class motor imagery	NA (forced error corrections)	2.3–7.6 letters/min
Common spatial patterns (vowel imagery) [‡]	Healthy	EEG	Direct two-vowel selection device	71% (68–78%)	NA (offline)
Sensorimotor rhythm (motor imagery)	Epileptic	ECoG	Two-class motor imagery	76% (64–88%)	0.41 letters/min (0.32–0.82 letters/min)
Neurotrophic Electrode: motor	Disabled (LIS)	Microelectrode	Cursor controlled AAC device	NA	3 letters/min
Neurotrophic Electrode: speech [‡]	Disabled (LIS)	Microelectrode	Continuous vowel synthesis device	45–70% (89% maximum)	0.57–7.97 bits/min [‡]

Performance accuracy and information transfer rates are taken from relevant works (citations provided in appropriate article sections).

When necessary, transfer rates are computed from reported values.

[‡]These methods perform direct speech prediction while others are indirect methods for letter spelling.

[‡]Only a single 'selection', in terms of end point production, is needed for the Neurotrophic Electrode: speech BCI.

AAC: Augmentative and alternative communication; ALS: Amyotrophic lateral sclerosis; BCI: Brain-computer interface; ECoG: Electroencephalography; EEG: Electroencephalography; LDA: Linear discriminant analysis; LIS: Locked-in syndrome; NA: Not applicable.