Cognitive-motor brain–machine interfaces

Ariel Tankus [a,b], Itzhak Fried [b,c,d], Shy Shoham [a,*]

[a] Faculty of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
[b] Functional Neurosurgery Unit, Tel-Aviv Medical Center, Tel-Aviv 64239, Israel
[c] Department of Neurosurgery, David Geffen School of Medicine, and Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA 90095, USA
[d] Sackler School of Medicine, Tel-Aviv University, Tel-Aviv 69978, Israel

ABSTRACT

Brain–machine interfaces (BMIs) open new horizons for the treatment of paralyzed persons, giving hope for the artificial restoration of lost physiological functions. Whereas BMI development has mainly focused on motor rehabilitation, recent studies have suggested that higher cognitive functions can also be deciphered from brain activity, bypassing low level planning and execution functions, and replacing them by computer-controlled effectors. This review describes the new generation of cognitive-motor BMIs, focusing on three BMI types:

1. Speech BMI – reconstructing a person's speech based on the neuronal activity.
2. Direct object control – controlling object movement without mimicking the limb movement that would yield the desired object movement.
3. Decoding internal processes, such as neuronal representations of sensory information and decision making.

By outlining recent progress in developing these BMI types, we aim to provide a unified view of contemporary research towards the replacement of behavioral outputs of cognitive processes by direct interaction with the brain.

© 2014 Published by Elsevier Ltd.

1. Introduction

Brain–machine interfaces (BMIs) are machines that can decode physiological signals from the brain and convert them into actions in a manner that reflects the brain's intention (Schwartz, 2004; Birbaumer, 2006; Leuthardt et al., 2006; Moran, 2010). So far, the vast majority of BMI research has focused on the restoration of movement functionality, bypassing paralyzed limbs, and replacing them with external effector such as robotic arms. However, a completely paralyzed person misses other functions as well, like the ability to speak for example, and their restoration can provide tremendous improvement in the patient's ability to communicate with their surroundings.

To undertake these challenges, the realm of brain–machine interfaces may be expanded by using signals recorded from various systems that were traditionally outside the central focus of the BMI field, including higher cognitive processes like working memory, attention, and mental processing networks (Jerbi et al., 2009). These processes, however, incorporate abilities unique to human beings, and therefore research results cannot generally be projected from nonhuman animals to the human, as is common in the motor literature (Georgopoulos et al., 1986; Taylor et al., 2002; Carmena et al., 2003; Paninski et al., 2004; Shoham et al., 2005). Thus, crucial data and insights into aspects of human cognition that may be harnessed for brain–machine interface research and applications relies on the rare opportunities to record in awake humans the intracortical electrical activity of single cells, and of small or large cellular assemblies; previously, such unique clinical opportunities were used to develop a basic-science understanding of neuronal encoding underlying human perception (Mukamel and Fried, 2012). Although these studies are typically performed with able-bodied subjects suffering from intractable epilepsy, movement or affective disorders, there are good reasons to expect that the observed neural encoding will share essential features with

volitional activations in paralyzed subjects (Shoham et al., 2001; Kokotilo et al., 2009).

This review focuses on the encoding and decoding of the activity of neuronal ensembles during processes which are on the border between cognitive processes (Andersen et al., 2010; Zander and Kothe, 2011) and motor processes (Hatsopoulos and Donoghue, 2009), and specifically, on the motor expressions of cognitive processes. We note that although functional magnetic resonance imaging (fMRI) can also provide important insights into the neuronal organization of these and related processes, and also has been used for internal process-decoding, it is left mostly outside the scope of this review, primarily because it only provides a very indirect measure of neuronal encoding and is impractical for a real-life BMI.

2. Speech brain–machine interfaces

Speech is perhaps the clearest example of a combination of cognitive and motor processes. Restoring the ability to (artificially) communicate again is a natural way for allowing paralyzed individuals to express their high level thoughts directly to other people, and can greatly improve the quality of life of ALS patients during the late stages of the disease. Indeed, much of BMI work has attempted to address this need indirectly: first by translating motor cortical volitions into binary commands (in the first BMI implanted in the human brain, (Kennedy and Bakay, 1998)) and later on using spelling devices for expressing thoughts in writing, typically controlled using a moving cursor (Birbaumer et al., 1999; Kennedy et al., 2000). Spelling devices can either be controlled noninvasively using either electroencephalography (EEG) (Birbaumer et al., 1999; Guger et al., 2009; Ryan et al., 2011), or fMRI (Sorger et al., 2012) or invasively, by using electrocorticography (ECoG), e.g., (Krusienski and Shih, 2011) or microelectrodes (Kennedy et al., 2000), see review in (Cecotti, 2011).

A more direct and probably much more natural and intuitive BMI strategy that has only recently been pursued and demonstrated is to directly decode speech and/or speech intentions from speech-related neural activity (for review see: (Brumberg and Guenther, 2010; Brumberg et al., 2010; Pei et al., 2012). Although the brain areas most commonly associated with speech-production are Broca's and face motor cortex areas, speech production involves a much larger network as described by multiple imaging, lesion and stimulation studies (Mohr, 1976; Ojemann and Mateer, 1979; Ghosh et al., 2008; Sahin et al., 2009; Hickok and Poeppel, 2007; Hickok, 2009; Hickok et al., 2011). The speech production network includes areas in charge of speech comprehension, for example, the bilateral superior temporal gyri (STG), inferior temporal cortex, left angular gyrus, pars orbitalis, and the bilateral superior temporal sulci (Price, 2010). The transition from percepts to actual motor output involves word retrieval (left middle frontal cortex) and articulatory planning (left anterior insula), with initiation and execution of speech controlled by the left putamen, pre-supplementary motor area (pre-SMA), SMA, and motor cortex (Price, 2010). When no actual motor output is produced, i.e. articulatory imagery, auditory-like responses still occur in auditory cortex, suggesting efference copies still predict the auditory outcome of imagined articulation (Tian & Poeppel, 2010). Imaging studies showed that both the rostral anterior cingulate cortex (Sahin et al., 2009; Paus et al., 1993) and the STG (Buchsbaum et al., 2001; Peeva et al., 2010) also participate in the basic control of speech production, while the medial-frontal orbitofrontal cortex is involved in speech comprehension (Sabri et al., 2008) and reading (Kujala et al., 2007). Anterior cingulate cortex is more active for speech than non-speech vocalization (Chang et al., 2009), and participates in speech motor control in both human (Wise et al., 1999;

Sörös et al., 2006) and monkey's vocalizations (Paus, 2001). Speech-related neural activity was characterized most carefully in Broca's area, the region most intimately linked to speech production and its loss in expressive (Broca's) aphasias, and where human studies have been used to characterize the temporal dynamics of speech-related local field potentials (Halgren et al., 1994; Sahin et al., 2009). Much less is known about the encoding of speech production at the single unit level. In a classical study, Ojemann et al. (1988) identified units in STG and adjacent superior margin of the middle temporal gyrus whose activity modulated with overt speech, supporting the area role in the motor aspects of speech, already inferred from the location of lesions producing permanent motor aphasias (Mohr, 1976).

As is usual in the BMI field, ECoG, local field potentials or single unit activity may be employed as signal sources. Although some studies show better performance when employing local field potentials or multiunit activity in comparison with spikes (for example: (Stark and Abeles, 2007)), most studies in the field presume that utilizing the information encoded by populations of single neurons can lead in principle to interfaces with higher information density. The first speech BMI was demonstrated in a paralyzed patient with a neurotrophic microelectrode in his precentral gyrus (Guenther et al., 2009). A Kalman filter-based decoder was trained to predict, from unit firing rates, the intended formant frequencies, which are the spectral peaks of the spectrum envelope amplitude of the sound (Fant, 1970). These frequencies were then synthesized to provide the subject with audio feedback. The data collected with this patient also served for decoding phonemes from unit firing rates; comparing three classification methods, linear discriminant analysis (LDA), support vector machine (SVM), and flexible discriminant analysis (FDA) – SVM achieved the highest accuracy (21%) for 38 phonemes (i.e., significantly above the chance level of 2.6%) (Brumberg et al., 2011).

In parallel to these early efforts with unit-based speech BMI, other researchers studied the ability to directly decode speech using ECoG – type signals. Kellis et al. (2010) recorded ECoG and micro-ECoG from the surface of face motor cortex and Wernicke's area in an epilepsy patient and classified a set of 10 spoken words on a trial-by-trial basis. Although high accuracy (90%) was achieved on average for all word pairs, it is difficult to compare these results to other studies, because the classified trials were not cross-validated and included only ones with subjectively-selected stereotyped vocal repetitions.

Schalk and colleagues showed that ECoG signals can be used to decode vowels and consonants in spoken or imagined words that could be employed for speech-based BMI systems (Pei et al., 2011). Combining time and frequency domain features, naïve Bayes classifiers obtained similar accuracies for both vowel decoding and consonant decoding tasks: around 40% accuracy for overt and covert speech tasks, significantly above the 25% chance level.

Schalk and colleagues have also demonstrated a cross-modality BMI where ECoG signals associated with different overt and imagined phoneme articulation were utilized for the control of a one-dimensional computer cursor rapidly and accurately. The phonetic content was distinguishable within higher gamma frequency oscillations and enabled users to achieve final target accuracies between 68% and 91% within 15 min. (Leuthardt et al., 2011). They suggested the use of speech-related signals as a cognitive and physiologic substrate for BMI operation, not limited to a speech-producing BMI.

Recently, we used single unit recordings obtained from human subjects during the pronunciation of speech segments, to propose a speech BMI that is based on direct decoding of the phoneme that the user wishes to pronounce (Tankus et al., 2012a) – decoded phonemes can be pre-recorded and played back. The decoder employs neurons recorded mainly from two populations of cells, each of