Journal of Physiology - Paris xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Physiology - Paris

journal homepage: www.elsevier.com/locate/jphysparis

Key considerations in designing a speech brain-computer interface

Florent Bocquelet [a,b], Thomas Hueber [c], Laurent Girin [c], Stéphan Chabardès [d], Blaise Yvert [a,b,*]

[a] INSERM, BrainTech Laboratory U1205, F-38000 Grenoble, France
[b] Univ. Grenoble Alpes, BrainTech Laboratory U1205, F-38000 Grenoble, France
[c] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France
[d] Grenoble University Hospital, F-38000 Grenoble, France

ARTICLE INFO

Article history:
Received 3 February 2017
Received in revised form 21 June 2017
Accepted 19 July 2017
Available online xxxx

Keywords:
Neural prosthesis
Neural decoding
Rehabilitation
Speech synthesis
Silent speech
Aphasia

ABSTRACT

Restoring communication in case of aphasia is a key challenge for neurotechnologies. To this end, brain-computer strategies can be envisioned to allow artificial speech synthesis from the continuous decoding of neural signals underlying speech imagination. Such speech brain-computer interfaces do not exist yet and their design should consider three key choices that need to be made: the choice of appropriate brain regions to record neural activity from, the choice of an appropriate recording technique, and the choice of a neural decoding scheme in association with an appropriate speech synthesis method. These key considerations are discussed here in light of (1) the current understanding of the functional neuroanatomy of cortical areas underlying overt and covert speech production, (2) the available literature making use of a variety of brain recording techniques to better characterize and address the challenge of decoding cortical speech signals, and (3) the different speech synthesis approaches that can be considered depending on the level of speech representation (phonetic, acoustic or articulatory) envisioned to be decoded at the core of a speech BCI paradigm.

1. Introduction

It is estimated that the prevalence of aphasia is about 0.3% of the population, which corresponds to more than 20 millions people worldwide. Such impairment occurs most often after a brain stroke, but this disability also affects people with severe tetraplegia consequently to an upper spinal cord trauma, locked-in individuals, people suffering from neuro or muscular degenerative diseases (such as amyotrophic lateral sclerosis (ALS), Parkinson's disease, or myopathies), and even comatose patients. For these people, speech loss is an additional affliction that worsens their condition: It makes the communication with caregivers very difficult, and more generally, it can lead to profound social isolation and even depression. Restoring communication abilities is thus crucial for these patients.

Different solutions for communication have been developed, most often consisting of word spelling devices making use of residual physiological signals, for example based on eye-tracking strategies possibly accompanied by a clicking capability. However, these solutions become inappropriate when people have lost too much of their motor functions. Communication systems controlled directly by brain signals have thus started to be developed to overcome this problem. This concept has been pioneered by Farwell and Donchin who proposed a spelling device based on the evoked potential P300 (Farwell and Donchin, 1988), a method that has since been used successfully by an ALS patient to communicate (Sellers et al., 2014). Other EEG-based approaches use steady-state potentials tuned at different frequencies (Middendorf et al., 2000). A great advantage of these approaches is their non-invasiveness. However, they have been limited by a low spelling speed of a few characters per minute, although recent improvements suggest that higher speeds could be achieved (Townsend and Platsko, 2016). Another major limitations of EEG-based BCI systems for communication is that they still require a high level of concentration of the subjects (Käthner et al., 2014; Baykara et al., 2016), imposing a high cognitive workload limiting their easy use over extensive periods of time. Interestingly, with the drawback to require invasive recordings, BCI systems based on intracortical signals seem to alleviate the subject fatigue, the external device becoming progressively embodied after a period of training (Hochberg et al., 2006, 2012; Collinger et al., 2013; Wodlinger et al., 2015). Recently, Jarosiewicz and colleagues showed that incorporating self-recalibrating algorithms into an intracortical brain-computer spelling interface allows spelling performances of about 20–30 characters per minute by people with severe paralysis over long periods of use (Jarosiewicz et al., 2015).

* Corresponding author at: INSERM, BrainTech Laboratory U1205, F-38000 Grenoble, France.
E-mail address: blaise.yvert@inserm.fr (B. Yvert).

http://dx.doi.org/10.1016/j.jphysparis.2017.07.002
0928-4257/© 2017 The Author(s). Published by Elsevier Ltd.

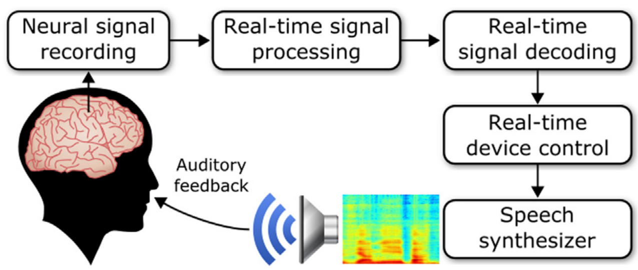

Fig. 1. Principle of a speech brain-computer interface.

The strategy of letter-selection BCI systems remains an indirect way of communicating based on movement direction decoded from the hand and/or arm area of the motor cortex. This is thus conceptually different than using speech, which is the natural and most efficient way of communication of the human species. Moreover, communication is often needed while other motor actions are performed requiring the resources of the hand/arm regions of the motor cortex (e.g. giving a phone call while moving in an environment or reaching for something). Thus, building a "speech BCI" to restore continuous speech directly from neural activity of speech-selective brain areas, as pioneered by Guenther and colleagues (Guenther et al., 2009), is an emerging field in which increasing efforts need to be invested in. As illustrated in Fig. 1, this strategy consists in extracting relevant neural signal features and converting them into input parameters for a speech synthesizer that runs in real time.

In this paper, we discuss several key requirements to restore speech with a BCI, including the choice of the speech cortical areas to record from, the recording techniques and decoding strategies that can be used, and finally the choice of speech synthesis approaches.

2. Choice of a brain region

Speech processing by the human brain involves a wide cortical network, which has been modeled by two main information streams linking auditory areas of the superior temporal plane to articulatory areas of frontal regions, one ventral and the other dorsal (Hickok and Poeppel, 2004, 2007). The ventral stream involves regions of the middle and inferior temporal lobe and maps speech sounds to meaning, while the dorsal stream runs through the dorsal part of the posterior temporal lobe at the temporo-parietal junction and is responsible for the sensori-motor integration of speech by mapping speech sounds to articulatory representations (Friederici, 2011; Hickok et al., 2011). Lesions of ventral stream regions of the temporal lobe result in Wernicke aphasia characterized by impairments of speech comprehension, while lesions of frontal areas result in Broca aphasia characterized by impairments of speech production. Classically, the dorsal stream has been described to be largely left-hemisphere dominant, but several studies indicate that many aspects of speech production activate cortical areas of the dorsal stream bilaterally (Pulvermüller et al., 2006; Peeva et al., 2010; Cogan et al., 2014; Geranmayeh et al., 2014; Keller and Kell, 2016).

Given this broad distribution of the speech network, to build a speech BCI, a choice needs to be made on the cortical areas to record and decode activity from. One possibility is to use signals from auditory areas of the ventral stream, which are known to encode the spectro-temporal representation of the acoustic content of speech, as assessed in both humans (Giraud et al., 2000; Formisano et al., 2008; Leonard and Chang, 2014; Leonard et al., 2015) and animals (Engineer et al., 2008; Mesgarani et al., 2008; Steinschneider et al., 2013). However, these areas are non-selectively involved in the sensory perception and integration of all speech sounds a person is exposed to. This includes self-produced speech but also other people speech, and even of non-speech environmental sounds as in the case for primary auditory areas. Thus, it can be expected that it would be difficult to identify activities specific to self speech intention in these areas. For this reason, probing neural activity in brain locations more specifically dedicated to speech production seems more relevant for conversational applications using a speech BCI (Guenther et al., 2009).

Several speech production conditions can be distinguished, including overt speech production, silent articulation (articulatory movements without vocalization, i.e. with no laryngeal activity), and inner (covert) speech production. The later condition (Perrone-Bertolotti et al., 2014) is likely the one most relevant when envisioning the use of a speech BCI by patients that intend to speak while not being able to produce articulatory movements. Articulatory speech production pathways originate from the speech motor cortex and project to the brainstem trigeminal, facial and ambiguus nuclei. Brainstem nuclei are difficult to access for recordings and there has yet been no evidence for their activation during covert intended speech. Thus, a speech BCI is likely to be easier to achieve by probing cortical areas underlying the production of inner speech.

Functional imaging studies have shown that overt word repetition activates motor and premotor cortices bilaterally (Petersen et al., 1988, 1989; Palmer et al., 2001; Peeva et al., 2010; Cogan et al., 2014). Continuous production of narrative speech was also shown to activate frontal motor speech regions and temporal and parietal areas bilaterally (Silbert et al., 2014). Intraoperative functional mapping data collected in a high number of patients undergoing awake surgery also report bilateral critical motor and premotor regions for overt speech production (Tate et al., 2014). The right hemisphere is also clearly activated during synchronized speaking in several regions including the temporal pole, inferior frontal gyrus, and supramarginal gyrus (Jasmin et al., 2016). When more complex tasks are considered that require additional semantic, lexical, or phonological processing, then specific activations are observed in the left inferior frontal cortex (Petersen et al., 1988, 1989; Price et al., 1994; Sörös et al., 2006; Basho et al., 2007). These findings suggest that speech production becomes left lateralized when inner high-level processing is required. In general, inner speech has been found to activate similar brain areas but with a lesser amplitude than overt speech across most ventral and dorsal stream areas (Price et al., 1994; Ryding et al., 1996; Palmer et al., 2001; Shuster and Lemieux, 2005). In particular, as for high-level overt speech production, cortical activity underlying covert speech production is left lateralized with strong activation of the left motor, premotor and inferior frontal cortex (Ryding et al., 1996; Palmer et al., 2001; Keller and Kell, 2016). The left inferior frontal cortex has further been shown to be specifically activated during covert word retrieval (Hirshorn and Thompson-Schill, 2006) and to be important for inner speech production as assessed using repetitive transcranial magnetic stimulation (Aziz-Zadeh et al., 2005). A careful anatomical voxel-based lesion study further confirmed the importance of this region as well as the white matter adjacent to the left supramarginal gyrus to achieve rhyme and homophone tasks requiring inner speech production (Geva et al., 2011).

Overall, the left inferior frontal region encompassing Broadman areas 4, 6, 44, 45, and 47, thus appears as a pertinent candidate from which to probe and decode neural activity for the control of a speech BCI. It should be noted that this strategy can only apply to aphasic patients whose speech networks remain intact, at least in this region. This is generally the case for instance for locked-in individuals or patients with ALS. To envision a speech BCI in people