

Towards a speech BCI:

The relation between sensorimotor cortex activity and
speech, articulator and facial movements

Efraim Salari

Towards a speech BCI: The relation between sensorimotor cortex activity and speech, articulator and facial movements

PhD Thesis, Utrecht University, The Netherlands

ISBN: 978-90-393-7106-0

Author: Efraïm Salari

Cover artwork & design: Efraïm Salari

Cover title: “Tussen de spleten in de muur groeide een nieuw idee”, naar de Loesje van vrijdag 17 augustus 2018.

Printed by: Ridderprint BV

Publication of this thesis was accomplished with financial support of the Brain Center Rudolf Magnus, Universty Medical Center Utrecht.

Copyright © 2019, Efraïm Salari

All rights reserved. No part of this thesis may be reproduced or transmitted in any form by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without the permission in writing of the copyright owner. The copyright of the articles that have been published or accepted for publication had been transferred to the respective journals.

**Towards a speech BCI:
The relation between sensorimotor cortex activity and
speech, articulator and facial movements**

*Op weg naar een spraak BCI:
De relatie tussen sensomotorische cortex activiteit en
spraak-, articulator- en gezichtsbewegingen*

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 1 april 2019 des middag te 4.15 uur.

door

Efraïm Salari

geboren op 20 maart 1991
te Tiel, Nederland

Promotor:

Prof. dr. N.F. Ramsey

Copromotoren:

Dr. Z.V. Freudenburg

Dr. M.J. van Steensel

*“I want to trip inside your head, spend the day there.
To hear the things you haven't said, and see what you might see.”*

Miracle drug - U2

Table of content

Introduction	1
Chapter 1 Spatial-Temporal Dynamics of the Sensorimotor Cortex: Sustained and Transient Activity	15
Chapter 2 Repeated Vowel Production Affects Features of Neural Activity In Sensorimotor Cortex	35
Chapter 3 The Influence of Prior Pronunciations on Sensorimotor Cortex Activity Patterns During Vowel Production	61
Chapter 4 Classification of Articulator Movements and Movement Direction From Sensorimotor Cortex Activity	81
Chapter 5 Classification of Facial Expressions for Emotional Communication Using BCI	105
Summary and Discussion	117
Conclusion	127
Nederlandse Samenvatting	129
Acknowledgements / Dankwoord	135
Scientific Achievements	141
About the Author	145



Introduction

*“I want to hear you when you call, do you feel anything at all?
I want to see your thoughts take shape, and walk right out.”*

Introduction

What Is A BCI And Why Is It Used?

Human communication involves the coordinated movement of many muscles. It has been claimed for instance that more than one hundred muscles act simultaneously to produce speech, and to place articulators, such as the tongue, lips, jaw and larynx, in the correct position (Levelt 1993; Meister et al. 2007; Guenther and Hickok 2015). Speech related movements need to follow each other with impressive velocities since people can speak about 15 sounds per second (Levelt 1993) and between 120-180 words per minute (Miller et al. 1976). In situations where speech is not possible, other forms of communication can be used that involve other forms of movements. When people for instance write letters, text a message or use sign language, they use hand and arm movements. Besides these types of language-based communication, communication can also be accomplished by body language, such as facial expressions to express emotions, or simple hand gestures that can convey much information. Taken together, in all forms of communication, movements are involved, being it verbal, non-verbal or both at the same time.

However, for people suffering from severe forms of paralysis, the ability to communicate their thoughts and feelings can be heavily affected. A situation in which an individual is cognitively intact but unable to move and communicate, is called locked-in syndrome (LIS; American Congress of Rehabilitation Medicine 1995). Patients suffering from this condition are essentially trapped (or locked) in their own body. LIS is usually caused by brainstem stroke or by neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS; Hayashi and Kato 1989; Smith and Delargy 2005). In the most severe cases, patients need to be ventilated through a tracheostoma and the only remaining movements are usually limited to those of the eye.

Over the past decades, assistive technology has been developed to help people with severe paralysis communicate their thoughts and regain some level of control over their environment. Examples of such technology are for instance eye-trackers which, by using infrared light, can locate where people are looking at on a computer screen filled with letters (Smith and Delargy 2005). This can be fed to smart technology that, subsequently, recognizes which letter(s) the patient wants to select. Spelled words and sentences can subsequently be pronounced through a speech computer. As such, eye-trackers and other types of Augmentative and Alternative Communication (AAC) technology that are controlled by residual movements, may contribute significantly to improve the quality of life of many people with LIS. In fact, it has been shown by using this technology, that people suffering from LIS often rate their life surprisingly positive (Posner et al. 2007; Rousseau et al. 2015). Importantly, however, there are situations in which muscle-based AAC tools do not work. For example, eye-trackers are difficult to use outside on sunny days (Vansteensel et al. 2016). Furthermore, some LIS patients do not have any remaining eye-movements (completely or total LIS; Smith and Delargy 2005). People suffering from ALS, for instance, may lose eye movements over time (Hayashi and Kato 1989), which makes the use of an eye-tracker for communication impossible. Even more striking maybe, there are also situations in which eye (or other) movements were never there to indicate if a patient is conscious or not. A study that used functional magnetic resonance imaging (fMRI) for instance, showed that some of the patients who were thought to be fully unconscious (in coma or vegetative state) were actually able to adjust their brain signals on

command (Monti et al. 2010). This indicates that there should be some form of consciousness in these patients, which they cannot express to the outside world as in these situations there are no remaining sources of muscle activity, which can be used for communication. In these situations, other methods to control AAC technology are needed.

Alternative technology to help these patients communicate is being developed. These systems are based solely on neural activity, bypassing muscle involvement, and are called Brain-Computer Interfaces (BCIs). BCIs record brain signals that patients have learned to control, and which subsequently can be converted into computer commands (Wolpaw et al. 2002). BCI technology enables patients to regain the ability to communicate. In recent years, the potential for such systems in assisting paralyzed patients with communication has been investigated, which showed that BCIs can be reliably used when other forms of communication fail (see for instance Vansteensel et al. 2016).

Brain Signal Recording Techniques For BCI

There are different forms of BCIs, using different recording techniques. The most commonly used and investigated BCIs use electrophysiological recordings. These can be divided into non-invasive recordings, such as electroencephalography (EEG), and invasive techniques such as epi- or subdural electrocorticography (ECoG) and needle or micro-array electrodes. Each of these recording techniques will be discussed below.

A. Non-Invasive Techniques

EEG records electrical potentials using electrodes placed on the scalp. Since the neurons in the neocortex are generally oriented in the same direction, electrical potentials of the neurons that fire at the same time add up, resulting in a large electrical potential (Buzsáki et al. 2012). While EEG has shown to be useful for reliably recording brain activity for BCI control (see below for examples), the fact that electrodes are placed on the head of a BCI user comes with certain disadvantages. First, this technique is quite sensitive to artefacts from for instance cable movements or residual facial movements, and the spatial resolution and signal-to-noise-ratio are worse than that of invasive recordings (Leuthardt et al. 2004; Buzsáki et al. 2012). In addition, the electrodes need to be placed on the subjects' head and recalibrated with every use, which usually involves the assistance of a researcher or trained caretaker. Third, wearing an EEG system may be inconvenient or uncomfortable in certain situations, for instance, outside when it rains or at night while sleeping, which makes this technique unsuitable in situations in which it is important to be able to communicate or call a caregiver at any time. Finally, the aesthetics of EEG systems is considered a disadvantage by many users. Since EEG equipment is very prominently visible for the outside world, some subjects may perceive to be distanced from other people, which plays a role in their willingness to use it (Blabe et al. 2015).

Other non-invasive measuring techniques, such as functional near-infrared spectroscopy (fNIRS), fMRI, and magnetoencephalography (MEG) have also been used in BCI research (e.g. Mellinger et al. 2007; Andersson et al. 2011; Sorger et al. 2012; Heger et al. 2014; Bleichner et al. 2015; Naseer and Hong 2015), but they are either not reliable enough for accurate BCI use or they require equipment that is not usable in a home setting or that is too expensive (Wolpaw et al. 2002).

However, some of these techniques, such as for instance fMRI, show high correspondence with recording techniques that are more practical for BCI use, such as ECoG (Logothetis et al. 2001; Hermes et al. 2012a; Siero et al. 2013, 2014), and they are frequently used to non-invasively explore neural activity for BCI control in large sets of subjects.

B. Invasive Techniques

An alternative to non-invasive recording techniques are implantable electrodes. These are electrodes that are inserted within the cortex tissue (i.e. needle electrodes) or placed on the surface of the brain (ECoG electrodes).

Needle electrodes can be used to record from single neurons or a small group of neurons (Buzsáki et al. 2012) and have been used very successfully to record neural activity for BCI in research settings with patients with severe paralysis (Hochberg et al. 2006, 2012; Truccolo et al. 2008; Collinger et al. 2013). Although very high levels of performance have been obtained with these needle electrodes, they have so far not been used in long-term home based BCI applications for patients. One of the reasons for this is that the stability of the signal deteriorates over time and may drop on the long term (Downey et al. 2018). It is unclear why the signal becomes unstable, but a possible reason could be that due to micromovements the signal is not recorded from the same neuron over time or that the neuron changes its firing rate. Furthermore, the technology for the required amplifiers to be fully implanted does not yet exist, and until recently, the reliability of BCI control was not yet good enough for unsupervised home use (although see for recent developments Milekovic et al. 2018). Alternatively, a BCI system using surface electrodes has been proven to be applicable for long-term accurate and stable BCI control with very high precision, even in an unsupervised home use setting (Vansteensel et al. 2016; Pels et al. in submission).

Surface ECoG electrodes measure electrical potentials from ensembles of neurons directly underneath the covered area (Manning et al. 2009; Miller 2010; Moran 2010; Buzsáki et al. 2012). They are used as a standard clinical assessment tool for localization of epileptogenic tissue in epilepsy patients who need brain surgery as part of their treatment. If patients agree, they may also take part in research experiments to learn more about brain functioning. Research with these patients over the last decades has been the basis for the development of long-term accurate BCI usage. In the University Medical Center Utrecht, some epilepsy patients have given permission to place an extra grid over a clinically non-interesting area, for research purposes only. These grids are usually smaller than the grids placed for clinical purposes and contain more and smaller electrodes with small inter-electrode distances (e.g. 3 mm compared to 10 mm in standard clinical grids). This enables more detailed registrations and it has been suggested that this might be beneficial for more accurate BCI applications (Ramsey et al. 2018).

Since invasive recording techniques require surgery, it is important to understand the willingness of LIS patients to use these techniques. Interestingly, the willingness of patients to undergo surgery seems to depend on the condition that people have (Lahr et al. 2015). For instance, since the condition of brain stem stroke patients is usually stable, they sometimes express that they would not want people to ‘mess’ with the one thing that is still working (their head/brain). Other patients, however, such as ALS patients, may see invasive BCI as a last resort as their condition usually deteriorates.

How to Control A Computer Through Brain Signals?

Multiple brain areas have been used as a source of BCI control signals. Below, I will discuss some of the most widely used and investigated BCI approaches.

A. Responsive BCIs

Responsive BCIs make use of the brains' response to external stimulation, examples of which are the visually evoked potential (VEP; e.g. Sutter 1992; Middendorf et al. 2000; Martinez et al. 2007) and the P300 potential (e.g. Donchin et al. 2000; Brunner et al. 2011). Visually evoked potentials can be measured from the visual areas of the brain and are neural responses to visual alterations. In a VEP speller application a screen is presented with multiple options that a BCI user can select (e.g. letters, buttons). These options alternate their visual appearance, each in a unique way (e.g. by flickering or changing color at a specific rate). Each of the selection options will result in a unique neural response and by analyzing the neural response amplitude or periodicity, it can be inferred which of the objects the user was looking at and wanted to select. The P300 potential is also elicited by visual alterations of a target but is usually recorded from parietal areas (Wolpaw et al. 2002), although also other areas have been used (see for instance Brunner et al. 2011). It is thought to reflect the response of the brain to the detection of an alteration in a significant and attended stimulus.

B. Self-Induced BCIs

Self-induced BCIs are not driven by external stimuli, but employ self-induced mental activity. Examples are for instance BCIs based on slow cortical potentials (SCP or readiness potentials; e.g. Birbaumer et al. 1999) or sensorimotor rhythms (Wolpaw et al. 2002). Slow cortical potentials are associated with the anticipation to a task and are recorded from the vertex as negative fluctuations in electrical potentials that arise between 0.5 seconds and several seconds after an event (Birbaumer et al. 1990). With training, people can control these potentials and use it for BCI control. Another widely used method for self-induced control of BCI is by recording activity related to (attempted) movement from the sensorimotor cortex (SMC). These recordings can be referred to as sensorimotor rhythms (Wolpaw et al. 2002).

The Sensorimotor Cortex

The sensorimotor cortex can be divided into two parts, the primary motor cortex (M1), which extends from the precentral sulcus into the central sulcus and the primary somatosensory cortex (S1), which is located between the central sulcus and the postcentral sulcus. Although it is widely recognized that the primary motor cortex is mainly responsible for motor execution and the primary somatosensory cortex for the processing of tactile information and proprioception (Penfield and Boldrey 1937; Kandel et al. 2000), the division between motor and sensory processing is not that clear cut and overlap in motor and somatosensory functioning has been found for both areas (Fetz et al. 1980; Soso and Fetz 1980). Together the primary motor and somatosensory areas can be referred to as the sensorimotor cortex and it has been shown that both M1 and S1 are tuned to specific body parts in a topographical (i.e. somatotopic) organization, meaning that adjacent subareas are tuned to adjacent body parts (Penfield and Boldrey 1937; Kandel et al. 2000; Dechent

and Frahm 2003; Schellekens et al. 2018). However, even though some areas and even single neurons may be mostly tuned to the control of a specific body part, they usually also show to be involved in the control of other body parts (Schieber 2001).

Sensorimotor Rhythms

Electrophysiological signals of brain activity are composed of different frequencies. An analogy can be drawn with a sound recording of a man with a low voice (producing low frequency oscillations in the sound signal) and a child with a high voice (producing high frequency oscillations in the sound signal), speaking at the same time. In a similar way, neurons can fire at different frequencies and if a signal is recorded from multiple neurons, the multiple frequencies can be extracted from that signal. Sensorimotor rhythms refer to specific frequencies within sensorimotor cortex electrophysiological recordings that show, compared to rest, a change in power during overt movements, attempted movements or imagined movements. It has been shown for instance that mu (8-12 Hz) and beta (18-26 Hz) rhythms decrease in power compared to baseline during movements and increase in power shortly after movement (Pfurtscheller and Lopes da Silva 1999). High-frequency band power has been indicated to increase compared to baseline during executed or attempted movement. It has been shown that these, self-induced, power changes can be used to accurately control a BCI (e.g. Vansteensel et al. 2016).

Speech BCI

One of the most intuitive concepts for controlling a BCI is that a BCI user simply thinks of a message he or she want to express, which the computer can subsequently recognize from the neural activity patterns and turn into textual or auditory output. Whereas the decoding of language itself may be rather difficult still, there is increasing interest in the use of the sensorimotor cortex for the decoding of movements related to speech (Kellis et al. 2010; Brumberg et al. 2011; Mugler et al. 2014; Herff et al. 2015; Ramsey et al. 2018). Most of these studies found that the sensorimotor cortex is a potential target for decoding brain activity related to speech. The rationale behind this is that each sound is formed by the movements of articulators such as the tongue, lips, jaw and larynx, which are known to be represented in the SMC (e.g. Penfield and Boldrey 1937; Grabski et al. 2012; Bouchard et al. 2013). Since each sound is accompanied with a unique set of articulator positions and movements (Alfonso and Baer 1982; Levelt 1993; Booij 1999; Rietveld and van Heuven 2001) and since movements of different parts of the body are known to be represented in different parts of the sensorimotor cortex (Penfield and Boldrey 1937; Grabski et al. 2012; Bouchard et al. 2013), it can be speculated that producing different sounds will result in different neural sensorimotor cortex patterns. These patterns indeed have been used for classification of different sounds based on neural activity. Previous studies have shown that phonemes, words, and even whole sentences can be classified using sensorimotor cortex activity recordings using ECoG or needle electrodes (Kellis et al. 2010; Brumberg et al. 2011; Herff et al. 2015; Ramsey et al. 2018; Stavisky et al. 2018).

Even though BCI systems are intended for people that cannot move, only a few studies investigated the sensorimotor cortex activity during covert/imagined speech (Pei et al. 2011a, b; Martin et al. 2014, 2016) or during attempted speech in someone who was paralyzed (Brumberg et

al. 2011). However, most studies investigated sensorimotor cortex activity during actual/overt speech (Blakely et al. 2008; Kellis et al. 2010; Leuthardt et al. 2011, 2013; Mugler et al. 2014; Herff et al. 2015; Ramsey et al. 2018). The reason for this is that with overt speech it is possible to relate what people said to the brain signals, while with covert or attempted speech this is more complicated. Importantly, it is thought that, compared to covert movements, the brain activity patterns associated with overt or executed movements can serve as a better proxy of the brain activity patterns during attempted movements. It has been shown, for instance, that different attempted hand movements (6 in total) performed by amputees can be distinguished from each other from the sensorimotor cortex activity without any form of somatosensory feedback (mean accuracy: $64\% \pm 14\%$; Bruurmijn et al. 2017) and that in tetraplegic patients the recognition accuracy of attempted finger movements (contrasted to rest) was significantly higher than that of imagined movements (Blokland et al. 2012). Furthermore, the activation of motor areas during imagined movements is only 30% of the activity seen during movement execution (Roth et al. 1996) or sometimes even absent (Hermes et al. 2011). This reduction in signal for imagined movements might be caused by inhibition processes involved in motor imagery: during covert movements, inhibition of neural activity may prevent people from actually making the movements (Guillot et al. 2012). Since neural data during executed movements by able-bodied participants can be more easily acquired compared to data of attempted movement by people with motor impairment, and since executed and attempted movements resemble each other in terms of neural activity patterns, experiments of overt speech are likely to be useful for the development of speech BCIs.

Signal Features

As stated above, the electrophysiological signal from the sensorimotor cortex contains several relevant frequency bands where movement-related changes occur. Studies on the most useful features for accurate speech classification based on neural signals showed that the use of high frequency band (HFB; >50 Hz) power changes resulted in the most accurate results of speech sound classification (Mugler et al. 2014). Indeed, many studies (Kellis et al. 2010; Leuthardt et al. 2011; Herff et al. 2015; Ramsey et al. 2018) have used HFB power signals to accurately classify speech units (phonemes, words, sentences). HFB power signals are thought to be related to the underlying firing of neurons (Manning et al. 2009; Miller et al. 2009; Miller 2010; Ray and Maunsell 2011) and have been found to be directly associated with movements of different articulators, mostly in the ventral sensorimotor cortex (Crone et al. 1998; Bouchard et al. 2013). Therefore, HFB power signals may be an optimal source signal for speech BCI.

These HFB power signals are most accurately recorded using invasive electrophysiological recordings (Leuthardt et al. 2004) and since previous studies have used these techniques for classification of speech sounds based on neural activity from the SMC, it is clear that there is a potential for speech BCIs based on sensorimotor cortex activity using HFB power recordings from ECoG surface or needle electrodes. However, there is up till now, not a speech BCI system used by patients at home. The reason for this might be that the accuracy values are still insufficient for application in daily life (around 20% accuracy for a complete set of American English phonemes both in attempted and executed pronunciation; Brumberg et al. 2011; Mugler et al. 2014) or because

the degrees of freedom are still limited (70% accuracy for 4 sounds and 75% for a set of 10 words; Herff et al. 2015; Ramsey et al. 2018). Since speech involves many different sounds, we have to find a balance between the ability to decode enough different sounds while at the same time remain high accuracy. I believe that, for the development of high-performance speech-BCIs (in terms of accuracy and degrees of freedom), it is important to better understand the brain signals related to speech, articulator and facial movements. Therefore, in this thesis I tried to get a better understanding of how the brain activity is related to these movements.

Thesis Outline

In this thesis, I focused on two questions. First, I wanted to know how variability in pronunciation of speech units (e.g. phonemes) relate to variability in sensorimotor cortex activity. The reason for this is that the manner in which people pronounce a particular sound and the context in which this happens (e.g. the surrounding sounds) may result in variable SMC activity patterns for that particular sound, which may affect the decodability of that sound based on neural activity. Variability is an important issue in decoding: the more similar the neural patterns of one sound are to each other and the more different they are compared to the patterns of other sounds, the easier it is for a computer algorithm to recognize and distinguish the neural activity patterns from each other. It is unknown however, how exactly variability in speech and articulator movements is related to variability of sensorimotor cortex activity, or how this is related to BCI decoding accuracy. To investigate this, I explored if and how variability in the production of elementary components of speech (single phonemes) is related to variability in sensorimotor cortex activity. I investigated three sources of variation that might be related to SMC activity variability, being (1) duration of utterance, (2) rate of utterance and (3) transition from one phoneme to another.

A. Speech Movements

In **chapter 1**, I investigated how variations in speech duration affect activity in the sensorimotor cortex. In order to do this, I developed methods for modelling sensorimotor cortex activity during different durations of speech and compared these to the actual recorded neural signals. In **chapter 2**, I investigated, if and how repeated production of the same sound is associated to consistent SMC neural activity profiles. This may not be straightforward since for finger movements it has been found that repeating the same movement may lead to different levels of neural activity (Hermes et al. 2012b). Therefore, I investigated this question for the more complex articulator movements that are involved in speech. In **chapter 3**, I investigated if the sensorimotor cortex neural activity related to the pronunciation of a sound is affected by previously produced sounds. The reason for this is that the pronunciation of a sound is known to be accompanied by slightly different articulator positions and movements depending on whether or not that sound was preceded by another sound and by the nature of the preceding sound. This effect is called coarticulation. Since the sensorimotor cortex is involved in the coordination of articulator positions and movements, it is likely that coarticulation effects are represented in the neural patterns of this area. It is important to investigate this effect because in natural speech a sound is rarely produced on its own but usually in the context of other sounds. In this chapter, I investigated how the existence of multiple activity patterns for

one sound may influence the accuracy scores of speech BCIs since it is unknown how coarticulation may influence speech BCI accuracy.

B. Non-Speech Movements

Besides the questions related to SMC activity variability related to actual speech, I also addressed a second question: can variations in movements of individual articulators and facial movements be distinguished from the neural activity patterns in the sensorimotor cortex, and if so, how? I investigated the neural activity patterns related to articulator and facial movements that are not related to speech, to get a better understanding of how SMC processes are related to individual as well as synchronized articulator and facial movements.

B.I. Articulator Movements

For articulator movements, we wanted to gain a better understanding of the neural patterns related to individual articulator movements. In **chapter 4**, I investigated at what spatial scale movements of different articulators can be distinguished from the sensorimotor cortex. In addition, I investigated if different movement directions of the same articulator, the tongue, can be distinguished from one another. The reason for this is that it is unclear where and at what spatial resolution movements of different articulators and different movements of the same articulator can be recognized based on the neural activity. This is an important question for communication BCIs as these movements may provide additional control signals to existing BCIs (e.g. the one described by Vansteensel et al. 2016), while keeping the size and risks of the required surgery minimal. In addition, these investigations may help to understand at what spatial scale speech related movements can be decoded.

B.II. Facial Movements

Finally, in **chapter 5**, I investigated how facial expressions are represented in the sensorimotor cortex and if they can also be distinguished from each other. Since emotional communication is an important feature of human communication, I think adding an emotional component to a speech communication BCI will add to the usage value of BCI systems. This is because facial expression movements reflect the emotional state someone wants to communicate. Facial expressions usually co-occur during speech and convey the emotional state that somebody wants to communicate. It is unclear, however, if facial expressions can be recognized from SMC activity. I used ECoG recordings to analyze SMC activity patterns related to facial expressions.

Summary and Discussion

Finally, I will give a summary of our findings and the results and their implications in light of existing literature. Furthermore, I will suggest future research directions.

References

- Alfonso PJ, Baer T (1982) Dynamics of Vowel Articulation. *Lang Speech* 25:151–173
- American Congress of Rehabilitation Medicine (1995) Recommendations for use of uniform nomenclature pertinent to patients with severe alterations in consciousness. *Arch Phys Med Rehabil* 76:205–209. doi: 10.1016/S0003-9993(95)80031-X
- Andersson P, Pluim JPW, Siero JCW, et al (2011) Real-Time Decoding of Brain Responses to Visuospatial Attention Using 7T fMRI. *PLOS ONE* 6:e27638. doi: 10.1371/journal.pone.0027638
- Birbaumer N, Elbert T, Canavan AG, Rockstroh B (1990) Slow potentials of the cerebral cortex and behavior. *Physiol Rev* 70:1–41. doi: 10.1152/physrev.1990.70.1.1
- Birbaumer N, Ghanayim N, Hinterberger T, et al (1999) A spelling device for the paralysed. *Nature* 398:297–298. doi: 10.1038/18581
- Blabe CH, Gilja V, Chestek CA, et al (2015) Assessment of brain-machine interfaces from the perspective of people with paralysis. *J Neural Eng* 12:043002. doi: 10.1088/1741-2560/12/4/043002
- Blakely T, Miller KJ, Rao RPN, et al (2008) Localization and classification of phonemes using high spatial resolution electrocorticography (ECoG) grids. In: 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008. EMBS 2008. pp 4964–4967
- Bleichner MG, Jansma JM, Salari E, et al (2015) Classification of mouth movements using 7 T fMRI. *J Neural Eng* 12:066026. doi: 10.1088/1741-2560/12/6/066026
- Blokland Y, Vlek R, Karaman B, et al (2012) Detection of event-related desynchronization during attempted and imagined movements in tetraplegics for brain switch control. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). pp 3967–3969
- Booij G (1999) *The Phonology of Dutch*. Clarendon Press
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–332. doi: 10.1038/nature11911
- Brumberg JS, Wright EJ, Andreassen DS, et al (2011) Classification of Intended Phoneme Production from Chronic Intracortical Microelectrode Recordings in Speech-Motor Cortex. *Front Neurosci* 5:65. doi: 10.3389/fnins.2011.00065
- Brunner P, Ritaccio AL, Emrich JF, et al (2011) Rapid Communication with a “P300” Matrix Speller Using Electrocorticographic Signals (ECoG). *Front Neurosci* 5. doi: 10.3389/fnins.2011.00005
- Bruurmijn MLCM, Pereboom IPL, Vansteensel MJ, et al (2017) Preservation of hand movement representation in the sensorimotor areas of amputees. *Brain* 140:3166–3178. doi: 10.1093/brain/awx274
- Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* 13:407–420. doi: 10.1038/nrn3241
- Camuñas-Mesa LA, Quiroga RQ (2013) A Detailed and Fast Model of Extracellular Recordings. *Neural Comput* 25:1191–1212. doi: 10.1162/NECO_a_00433
- Cheney PD, Fetz EE (1980) Functional classes of primate corticomotoneuronal cells and their relation to active force. *J Neurophysiol* 44:773–791
- Collinger JL, Wodlinger B, Downey JE, et al (2013) High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet* 381:557–564. doi: 10.1016/S0140-6736(12)61816-9
- Crone NE, Miglioretti DL, Gordon B, Lesser RP (1998) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* 121:2301–2315. doi: 10.1093/brain/121.12.2301
- Davis J (ed) (2015) *Phonetics: fundamentals, potential applications and role in communicative disorders*. Nova Publishers, New York
- Dechent P, Frahm J (2003) Functional somatotopy of finger representations in human primary motor cortex. *Hum Brain Mapp* 18:272–283. doi: 10.1002/hbm.10084
- Donchin E, Spencer KM, Wijesinghe R (2000) The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Trans Rehabil Eng* 8:174–179. doi: 10.1109/86.847808
- Downey JE, Schwed N, Chase SM, et al (2018) Intracortical recording stability in human brain-computer interface

- users. *J Neural Eng* 15:046016. doi: 10.1088/1741-2552/aab7a0
- Fetz EE, Finocchio DV, Baker MA, Soso MJ (1980) Sensory and motor responses of precentral cortex cells during comparable passive and active joint movements. *J Neurophysiol* 43:1070–1089
- Grabski K, Lamalle L, Vilain C, et al (2012) Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Hum Brain Mapp* 33:2306–2321. doi: 10.1002/hbm.21363
- Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14–23. doi: 10.1016/j.tics.2005.11.006
- Guenther FH, Hickok G (2015) Chapter 9 - Role of the auditory system in speech production. In: Aminoff MJ, Boller F, Swaab DF (eds) *Handbook of Clinical Neurology*. Elsevier, pp 161–175
- Guillot A, Di Rienzo F, MacIntyre T, et al (2012) Imagining is Not Doing but Involves Specific Motor Commands: A Review of Experimental Data Related to Motor Inhibition. *Front Hum Neurosci* 6:. doi: 10.3389/fnhum.2012.00247
- Hayashi H, Kato S (1989) Total manifestations of amyotrophic lateral sclerosis: ALS in the totally locked-in state. *J Neurol Sci* 93:19–35. doi: 10.1016/0022-510X(89)90158-5
- Heger D, Herff C, Putze F, et al (2014) Continuous affective states recognition using functional near infrared spectroscopy. *Brain-Comput Interfaces* 1:113–125. doi: 10.1080/2326263X.2014.912884
- Herff C, Heger D, de Pestors A, et al (2015) Brain-to-text: decoding spoken phrases from phone representations in the brain. *Neural Technol* 9:217. doi: 10.3389/fnins.2015.00217
- Hermes D, Miller KJ, Vansteensel MJ, et al (2012a) Neurophysiologic correlates of fMRI in human motor cortex. *Hum Brain Mapp* 33:1689–1699. doi: 10.1002/hbm.21314
- Hermes D, Siero JCW, Aarnoutse EJ, et al (2012b) Dissociation between Neuronal Activity in Sensorimotor Cortex and Hand Movement Revealed as a Function of Movement Rate. *J Neurosci* 32:9736–9744. doi: 10.1523/JNEUROSCI.0357-12.2012
- Hermes D, Vansteensel MJ, Albers AM, et al (2011) Functional MRI-based identification of brain areas involved in motor imagery for implantable brain-computer interfaces. *J Neural Eng* 8:025007. doi: 10.1088/1741-2560/8/2/025007
- Hochberg LR, Bacher D, Jarosiewicz B, et al (2012) Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature* 485:372–375. doi: 10.1038/nature11076
- Hochberg LR, Serruya MD, Friehs GM, et al (2006) Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 442:164–171. doi: 10.1038/nature04970
- Kandel E, Schwartz J, Jessell T (2000) *Principles of Neural Science*. McGraw-Hill Medical
- Kellis S, Miller K, Thomson K, et al (2010) Decoding spoken words using local field potentials recorded from the cortical surface. *J Neural Eng* 7:056007. doi: 10.1088/1741-2560/7/5/056007
- Lahr J, Schwartz C, Heimbach B, et al (2015) Invasive brain-machine interfaces: a survey of paralyzed patients' attitudes, knowledge and methods of information retrieval. *J Neural Eng* 12:043001. doi: 10.1088/1741-2560/12/4/043001
- Lee KF (1990) Context-independent phonetic hidden Markov models for speaker-independent continuous speech recognition. *IEEE Trans Acoust Speech Signal Process* 38:599–609. doi: 10.1109/29.52701
- Leuthardt EC, Cunningham J, Barbour D (2013) Towards a Speech BCI Using ECoG. In: Guger C, Allison BZ, Edlinger G (eds) *Brain-Computer Interface Research*. Springer Berlin Heidelberg, pp 93–110
- Leuthardt EC, Gaona C, Sharma M, et al (2011) Using the electrocorticographic speech network to control a brain-computer interface in humans. *J Neural Eng* 8:036004. doi: 10.1088/1741-2560/8/3/036004
- Leuthardt EC, Schalk G, Wolpaw JR, et al (2004) A brain-computer interface using electrocorticographic signals in humans. *J Neural Eng* 1:63. doi: 10.1088/1741-2560/1/2/001
- Levelt WJM (1993) *Speaking: From Intention to Articulation*. MIT Press
- Logothetis NK, Pauls J, Augath M, et al (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157. doi: 10.1038/35084005
- Manning JR, Jacobs J, Fried I, Kahana MJ (2009) Broadband shifts in LFP power spectra are correlated with single-neuron spiking in humans. *J Neurosci Off J Soc Neurosci* 29:13613. doi: 10.1523/JNEUROSCI.2041-09.2009
- Martin S, Brunner P, Holdgraf C, et al (2014) Decoding spectrotemporal features of overt and covert speech from

- the human cortex. *Front Neuroengineering* 7. doi: 10.3389/fneng.2014.00014
- Martin S, Brunner P, Iturrate I, et al (2016) Word pair classification during imagined speech using direct brain recordings. *Sci Rep* 6:25803. doi: 10.1038/srep25803
- Martinez P, Bakardjian H, Cichocki A (2007) Fully Online Multicommand Brain-Computer Interface with Visual Neurofeedback Using SSVEP Paradigm. In: *Comput. Intell. Neurosci.*
- Meister IG, Wilson SM, Deblieck C, et al (2007) The Essential Role of Premotor Cortex in Speech Perception. *Curr Biol* 17:1692–1696. doi: 10.1016/j.cub.2007.08.064
- Mellinger J, Schalk G, Braun C, et al (2007) An MEG-based Brain-Computer Interface (BCI). *NeuroImage* 36:581–593. doi: 10.1016/j.neuroimage.2007.03.019
- Middendorf M, McMillan G, Calhoun G, Jones KS (2000) Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Trans Rehabil Eng* 8:211–214. doi: 10.1109/86.847819
- Milekovic T, Sarma AA, Bacher D, et al (2018) Stable long-term BCI-enabled communication in ALS and locked-in syndrome using LFP signals. *J Neurophysiol* 120:343–360. doi: 10.1152/jn.00493.2017
- Miller KJ (2010) Broadband Spectral Change: Evidence for a Macroscale Correlate of Population Firing Rate? *J Neurosci* 30:6477–6479. doi: 10.1523/JNEUROSCI.6401-09.2010
- Miller KJ, Sorensen LB, Ojemann JG, Nijs M den (2009) Power-Law Scaling in the Brain Surface Electric Potential. *PLOS Comput Biol* 5:e1000609. doi: 10.1371/journal.pcbi.1000609
- Miller N, Maruyama G, Beaber RJ, Valone K (1976) Speed of speech and persuasion. *J Pers Soc Psychol* 34:615–624. doi: 10.1037/0022-3514.34.4.615
- Monti MM, Vanhaudenhuyse A, Coleman MR, et al (2010) Willful Modulation of Brain Activity in Disorders of Consciousness. *N Engl J Med* 362:579–589. doi: 10.1056/NEJMoa0905370
- Moran D (2010) Evolution of brain–computer interface: action potentials, local field potentials and electrocorticograms. *Curr Opin Neurobiol* 20:741–745. doi: 10.1016/j.conb.2010.09.010
- Mugler EM, Patton JL, Flint RD, et al (2014) Direct classification of all American English phonemes using signals from functional speech motor cortex. *J Neural Eng* 11:035015. doi: 10.1088/1741-2560/11/3/035015
- Murray GM, Sessle BJ (1992) Functional properties of single neurons in the face primary motor cortex of the primate. III. Relations with different directions of trained tongue protrusion. *J Neurophysiol* 67:775–785. doi: 10.1152/jn.1992.67.3.775
- Naseer N, Hong K-S (2015) fNIRS-based brain-computer interfaces: a review. *Front Hum Neurosci* 9. doi: 10.3389/fnhum.2015.00003
- Okada K, Matchin W, Hickok G (2018) Phonological Feature Repetition Suppression in the Left Inferior Frontal Gyrus. *J Cogn Neurosci* 1–9. doi: 10.1162/jocn_a_01287
- Pei X, Barbour DL, Leuthardt EC, Schalk G (2011a) Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *J Neural Eng* 8:046028. doi: 10.1088/1741-2560/8/4/046028
- Pei X, Leuthardt EC, Gaona CM, et al (2011b) Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *NeuroImage* 54:2960–2972. doi: 10.1016/j.neuroimage.2010.10.029
- Penfield W, Boldrey E (1937) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain J Neurol* 60:389–443. doi: 10.1093/brain/60.4.389
- Pfurtscheller G, Lopes da Silva FH (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin Neurophysiol* 110:1842–1857. doi: 10.1016/S1388-2457(99)00141-8
- Plomp R (2001) *The Intelligent Ear: On the Nature of Sound Perception*. Psychology Press
- Posner JB, Plum F, Saper CB, Schiff N (2007) *Plum and Posner’s Diagnosis of Stupor and Coma*. Oxford University Press, USA
- Ramsey NF, Salari E, Aarnoutse EJ, et al (2018) Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage*. doi: 10.1016/j.neuroimage.2017.10.011
- Ray S, Maunsell JHR (2011) Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLOS Biol* 9:e1000610. doi: 10.1371/journal.pbio.1000610
- Rietveld A, van Heuven V (2001) *Algemene Fonetiek*, 2nd edn. Coutinho, Bussum

- Roth M, Decety J, Raybaudi M, et al (1996) Possible involvement of primary motor cortex in mentally simulated movement: a functional magnetic resonance imaging study. *Neuroreport* 7:1280–1284
- Rousseau M-C, Baumstarck K, Alessandrini M, et al (2015) Quality of life in patients with locked-in syndrome: Evolution over a 6-year period. *Orphanet J Rare Dis* 10:88. doi: 10.1186/s13023-015-0304-z
- Saarinen T, Laaksonen H, Parviainen T, Salmelin R (2006) Motor Cortex Dynamics in Visuomotor Production of Speech and Non-speech Mouth Movements. *Cereb Cortex* 16:212–222. doi: 10.1093/cercor/bhi099
- Schellekens W, Petridou N, Ramsey NF (2018) Detailed somatotopy in primary motor and somatosensory cortex revealed by Gaussian population receptive fields. *NeuroImage* 179:337–347. doi: 10.1016/j.neuroimage.2018.06.062
- Schieber MH (2001) Constraints on Somatotopic Organization in the Primary Motor Cortex. *J Neurophysiol* 86:2125–2143
- Siero JC, Hermes D, Hoogduin H, et al (2013) BOLD Consistently Matches Electrophysiology in Human Sensorimotor Cortex at Increasing Movement Rates: A Combined 7T fMRI and ECoG Study on Neurovascular Coupling. *J Cereb Blood Flow Metab* 33:1448–1456. doi: 10.1038/jcbfm.2013.97
- Siero JC, Hermes D, Hoogduin H, et al (2014) BOLD matches neuronal activity at the mm scale: A combined 7 T fMRI and ECoG study in human sensorimotor cortex. *NeuroImage* 101:177–184. doi: 10.1016/j.neuroimage.2014.07.002
- Smith E, Delargy M (2005) Locked-in syndrome. *BMJ* 330:406–409. doi: 10.1136/bmj.330.7488.406
- Sorger B, Reithler J, Dahmen B, Goebel R (2012) A real-time fMRI-based spelling device immediately enabling robust motor-independent communication. *Curr Biol* CB 22:1333–1338. doi: 10.1016/j.cub.2012.05.022
- Soso MJ, Fetz EE (1980) Responses of identified cells in postcentral cortex of awake monkeys during comparable active and passive joint movements. *J Neurophysiol* 43:1090–1110
- Stavisky S, Rezaei P, Willett F, et al (2018) Decoding Speech from Intracortical Multielectrode Arrays in Dorsal “Arm/Hand Areas” of Human Motor Cortex.
- Sutter EE (1992) The brain response interface: communication through visually-induced electrical brain responses. *J Microcomput Appl* 15:31–45. doi: 10.1016/0745-7138(92)90045-7
- Truccolo W, Friehs GM, Donoghue JP, Hochberg LR (2008) Primary Motor Cortex Tuning to Intended Movement Kinematics in Humans with Tetraplegia. *J Neurosci* 28:1163–1178. doi: 10.1523/JNEUROSCI.4415-07.2008
- Vansteensel MJ, Pels EGM, Bleichner MG, et al (2016) Fully Implanted Brain–Computer Interface in a Locked-In Patient with ALS. *N Engl J Med* 375:2060–6. doi: 10.1056/NEJMoa1608085
- Wolpaw JR, Birbaumer N, McFarland DJ, et al (2002) Brain–computer interfaces for communication and control. *Clin Neurophysiol* 113:767–791. doi: 10.1016/S1388-2457(02)00057-3
- Yao D, Yamamura K, Narita N, et al (2002) Neuronal Activity Patterns in Primate Primary Motor Cortex Related to Trained or Semiautomatic Jaw and Tongue Movements. *J Neurophysiol* 87:2531–2541. doi: 10.1152/jn.2002.87.5.2531
- Zhang D, Song H, Xu R, et al (2013) Toward a minimally invasive brain–computer interface using a single subdural channel: A visual speller study. *NeuroImage* 71:30–41. doi: 10.1016/j.neuroimage.2012.12.069





Chapter 1

*“Freedom has a scent
Like the top of a newborn baby’s head.”*

Chapter 1

Spatial-Temporal Dynamics of the Sensorimotor Cortex: Sustained and Transient Activity

E. Salari, Z.V. Freudenburg, M.J. Vansteensel, N.F. Ramsey

Published Article: © 2018 IEEE. Reprinted, with permission, from E. Salari, Z.V. Freudenburg, M.J. Vansteensel, N.F. Ramsey, Spatial-Temporal Dynamics of the Sensorimotor Cortex: Sustained and Transient Activity, Transactions on Neural Systems and Rehabilitation Engineering, March 2018

Abstract

Background: How the sensorimotor cortex is organized with respect to controlling different features of movement is unclear. One unresolved question concerns the relation between the duration of an action and the duration of the associated neuronal activity change in the sensorimotor cortex. **Methods:** Using subdural electrocorticography electrodes, we investigated in five subjects, whether high frequency band (HFB; 75-135 Hz) power changes have a transient or sustained relation to speech duration, during pronunciation of the Dutch /i/ vowel with different durations. **Results:** We showed that the neuronal activity patterns recorded from the sensorimotor cortex can be directly related to action duration in some locations, whereas in other locations, during the same action, neuronal activity is transient, with a peak in HFB activity at movement onset and/or offset. **Discussion:** This data sheds light on the neural underpinnings of motor actions and we discuss the possible mechanisms underlying these different response types.

Keywords; Sensorimotor Cortex, Movement, Duration, ECoG, Speech

Introduction

Voluntary body movements are controlled by the sensorimotor areas of the brain (Kandel et al. 2000; Taylor and Gross 2003). Studies on the relation between sensorimotor cortex activity and body movements have revealed, among others, that the sensorimotor cortex is somatotopically organized. This somatotopic organization seems quite detailed, since individual fingers (Dechent and Frahm 2003; Miller et al. 2009b; Siero et al. 2014) and even separate muscles (Hadoush et al. 2011; Espadaler et al. 2012) can be distinguished from one another within this brain area. Besides a topographical organization, there is evidence from both monkey and human studies that the sensorimotor cortex has a role in controlling different features of a movement. It has been shown for instance that the activity of single sensorimotor neurons can be related to the position of a body part (Wang et al. 2007; Truccolo et al. 2008), but also to movement direction (Georgopoulos et al. 1982; Moran and Schwartz 1999) or velocity (Truccolo et al. 2008; Wang et al. 2007; Moran and Schwartz 1999), change in force (Georgopoulos et al. 1992) or to motor planning (Tanji and Evarts 1976; Donoghue et al. 1998). How the sensorimotor cortex is organized with respect to controlling these different movement features remains to be determined. Importantly, the relationship between neural activity and movement features may not be straightforward, since there is evidence that repeating the same movement shortly after one another is not accompanied by the same magnitude of neuronal signal change for every repetition, despite equal behavioural output (Hermes et al. 2012). This indicates that neuronal activity is not always linearly related to movement and can depend on past actions. Clearly, the role of the sensorimotor cortex in controlling body movements is complex and the underlying mechanisms are not yet completely understood. Investigating the detailed relationship between the spatial-temporal neural patterns within this area and overt body movements may contribute to our understanding of the functioning of this area, which is also of importance for neural engineering purposes such as the development of brain-computer interface (BCI) systems.

One unresolved question involves the relation between the duration of neuronal activity and the duration of a motor action. In other words, is neuronal activity continuously or transiently related to motor output? Evidence for the existence of both transient and sustained responses mainly comes from single cell primate studies (Tanji and Evarts 1976; Fetz et al. 1980; Soso and Fetz 1980; Cheney and Fetz 1980; Donoghue et al. 1998). The topic received relatively little attention in studies with human subjects, but evidence for both transient and sustained responses have also been found in humans (Crone et al. 1998; Ball et al. 2008; Conant et al. 2018) and some studies have shown an effect of duration on the neural response profiles (Ball et al. 2008; Conant et al. 2018; Wang et al. 2014, 2017). It remains to be determined, however, how exactly the duration of movement is encoded in the brain with respect to these two types of responses, and whether or not extended action duration is associated with a corresponding temporal extension of neuronal activity. Furthermore, a detailed spatial mapping of both sustained and transient responses in humans is still missing for complex movements such as those involved in speech.

In the current study, we investigated the relation between sensorimotor cortex activity duration and action duration. Since people can quite easily vary the duration of the pronunciation of vowels, and thereby the duration of a motor action, we focused on the motor cortex areas

involved in articulator movements and speech pronunciation. Articulator movements are known to be controlled by the ventral parts of the sensorimotor cortex (Crone et al. 2001; Towle et al. 2008; Pei et al. 2011; Bouchard et al. 2013) with the larynx and tongue being represented more ventrally and the jaw and lips more dorsally (Bouchard et al. 2013). Indeed, it has been demonstrated that articulator movements (Bleichner et al. 2015), as well as speech units (Kellis et al. 2010; Mugler et al. 2014) can be distinguished ('classified') reliably from this area.

We recorded neural signals with electrocorticography (ECoG), a technique that benefits from a unique combination of high temporal resolution, comparable to electroencephalography (EEG), and high cortical sampling specificity (Siero et al. 2014). Using this technique, it has been demonstrated that movements of, for instance, the tongue, lips, hand or foot, are accompanied by an increase in high-frequency-band (HFB; >50 Hz) power in the sensorimotor cortex (Crone et al. 1998; Bouchard et al. 2013; Miller et al. 2007). This HFB power increase is thought to be associated with underlying neuronal firing (Manning et al. 2009; Miller et al. 2009a; Ray and Maunsell 2011). Therefore, we used HFB power changes associated with sustained pronunciation of single vowels to investigate whether neural activity has a transient or sustained relation to speech duration.

Methods

Participants

Five subjects (age 14-41y; median 21y, 3 females), who were implanted with subdural ECoG electrodes for the treatment of epilepsy in the University Medical Center Utrecht participated in this study. Three subjects (A, B & E) had coverage with standard clinical grids (exposed electrode diameter of 2.3 mm with a 10 mm inter-electrode distance) over the ventral sensorimotor cortex (vSMC; left hemisphere in 2 subjects, right in 1 subject). Two subjects (C & D) gave permission to place, besides the clinical grids, an extra high-density (HD) electrode grid (exposed electrode diameter 1 mm for subject B and 1.17 mm for subject C, inter-electrode distance 4 mm for both subjects) over the left mouth sensorimotor cortex for research purposes. For these subjects, only the HD electrodes were used for the current analysis.

This study was approved by the ethics committee of the University Medical Center Utrecht and is in accordance with the Declaration of Helsinki (2013). All subjects gave written informed consent.

Task

Subjects were asked to perform a vowel durations task in which they pronounced the Dutch /i/ vowel for 1, 2 or 3 seconds. This phoneme was chosen as it is easy to pronounce for variable durations and because it engages multiple articulators, including the tongue (Booij 1999; Rietveld and van Heuven 2001), which is well represented in sensorimotor cortex (Penfield and Boldrey 1937; Bouchard et al. 2013). The task was presented on a computer screen, which was placed at a distance of about 1 m. A trial started with a 500 ms cue indicating the pronunciation duration (1, 2 or 3 seconds) to prepare subjects. After 1000 ms, a visual cue (presented by the letters 'ie', corresponding, in Dutch, to the /i/ sound from the international phonetic alphabet) instructed the

subjects to start the pronunciation and hold it for as long as the visual cue was visible, which was followed by an inter-trial interval (fixation cross presented) of 2000 ms. Trials were randomized and each type was repeated 15 times per run (3 subjects performed two runs).

Data Acquisition

Signals were recorded at a sampling rate of 512Hz for subjects A, B, C (run 2), and E, at 2048Hz for subject C (run 1; all Micromed, Treviso, Italy) and at 2000Hz for subject D (Blackrock Microsystems LLC, Salt Lake City, USA). Three subjects performed the task twice (A, C & D), and for these subjects the data of the two runs were concatenated (for subject C the data of the first run was down-sampled to 512Hz for concatenation). Vowel pronunciation was recorded using microphones installed in the patient's room. All data was processed and analyzed using Matlab software (The Mathworks, Inc., Natick, MA, USA), unless specified otherwise.

Data Preprocessing

First, deviations in the power-density distributions, line noise values and raw voltage distributions, some of which are described in Liu et al. (2015), were used to identify electrodes with noisy or flat signals, which were subsequently removed from further analysis. Electrodes that were classified as ictal by a neurologist were also removed from further analysis. For the remaining electrodes, line noise and harmonics thereof were removed using a 3rd order butterworth filter ('butter' function and 'filtfilt' function from Matlab) and a common average re-reference was performed. Finally, signals were visually inspected and any trials with excessive noise (which sometimes occurs as a result of cable movements) were removed from the analysis.

Subsequently, electrode positions were visualized on the 3D rendering of the pre-surgical MRI scan using an in-house developed procedure (Branco et al. 2016; Hermes et al. 2010). Electrodes that were located over the sensorimotor cortex (pre- and postcentral gyrus) as indicated by the freesurfer MRI segmentation (Fischl 2012) were identified by visual inspection. All further analyses were performed using only these electrodes (25, 18, 38, 114 and 8 electrodes for subjects A-E respectively), excluding the identified noisy and ictal electrodes.

For every retained sensorimotor cortex electrode, the high frequency band (75-135 Hz) power was calculated per sample point by applying a Gabor wavelet function (Bruns 2004) for all frequencies in the HFB range in bins of 1 Hz, with a full width half maximum (fwhm) of 4 wavelets per frequency, followed by a log transformation ($10 \cdot \log_{10}$) and averaging (over frequencies between 75 and 135Hz) of the resulting values. HFB power values (for all time points) were subsequently normalized by subtracting the mean signal value of the whole time series and dividing the result by the standard deviation of the whole time series. Finally, this signal was smoothed with a moving average window of 500 ms (window centered around each sample). We chose the 500 ms window because we found this to be the optimal window for (speech) movement classification in a previous study (Branco et al. 2018). In that study the smoothing window was determined using an elaborate optimization algorithm for several parameters, using classification of spatiotemporal HFB patterns as optimization outcome measure. The current frequency band (75-135Hz) was chosen such that the lower bound matches that of previous experiments (Crone et al. 1998). The

upper bound was determined by hardware filters. The audio signal was aligned with the brain signal and the voice onsets and offsets were automatically determined by a vowel detection algorithm (Hermes 1990) and subsequently checked and corrected using Praat annotation software (Boersma 2002). Deviations in voice onset and offset from the cued timing were corrected in the brain signal by interpolation or down sampling, so that the brain signals of different trials could be adequately aligned. This step was necessary to allow for comparison of different (complete) trials, including both voice onset and offset. Only by making sure that each trial has the same duration, can complete trials be averaged and visualized as a whole, including a reliable, undiluted estimation of the neural responses associated with both voice onset and offset and the period in between. Visual inspection of the original and corrected time-series did not reveal any major differences other than the expected slight timing differences, indicating that the interpolation and down-sampling did not have a major influence on the signal values.

Statistical Procedures

A regression analysis with three predictors was performed on the whole (corrected) time series to describe the electrode responses to the task (Figure 1) and to determine which electrodes responded significantly to the task. The predictors consisted of a transient response around voice onset, a sustained response between voice on- and offset, and a transient response around voice offset. The first and third predictor were created by convolving a Gaussian function to an impulse function with impulses at voice onset and offset respectively (corrected such that the Gaussian peak would be at voice onset or offset). The width of the Gaussian (fwhm) was estimated for every subject separately (mean = 0.65 seconds, range = 0.51 - 0.76 seconds) using data from a different task. Similar to the vowel durations task, subjects were required to repeatedly pronounce the same vowel (*/i/*), but in this case very briefly for three times at a 1 Hz repetition rate per trial. For this task, we determined the significant electrodes by doing a simple r^2 -analysis, contrasting speech versus silence periods. We aligned all trials (13 per run, subjects A, C & D performed the task twice) to the first pronunciation and calculated the fwhm of the average peak (over trials and significant electrodes) in HFB power related to the first pronunciation. Note that, although potentially the ‘shape’ of a (transient) response can be different for different electrodes and therefore there doesn’t necessarily have to be one standard response, inspection of the data showed that this mean response described the data well and could be used for finding significant electrodes. The second predictor was created in a similar way as the transient predictors but with a block function between the impulses of the first and third predictor. Since the task had trials of 1, 2 or 3 second duration, we could evaluate whether sustained responses showed a relation with action duration.

The onset of movement-related activity can be different for different locations in the brain. This effect of shifting activity onset has been reported before in both monkey and human studies (Hermes et al. 2012; Crone et al. 1998; Bouchard et al. 2013; Coudé et al. 2011) and has for speech been attributed to the involvement of different articulators (Bouchard et al. 2013). We repeated the regression analysis procedure while shifting all predictors in steps of 0.1 second, from 0.5 seconds before voice onset/offset to 0.5 seconds after, to capture activity related to the movement as accurate as possible.

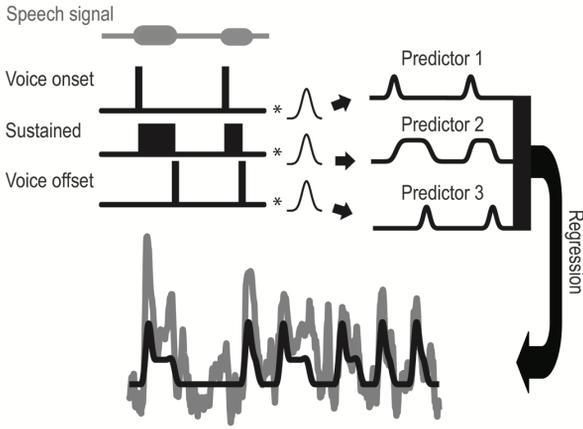


Figure 1. The regression procedure with the used predictors for the whole time series. The first and third predictor (right panel) were associated with voice onset and offset and were created by convolving an impulse function (left panel) at voice onsets and voice offsets respectively, with a subject specific Gaussian function over the whole time series. The second predictor, which was related to sustained activity during vowel pronunciation, was created by convolving a block function (left) between voice onset and voice offset with the same Gaussian function. Horizontal bars at the top (speech signal) indicate the time periods where subjects spoke. A regression was performed and the signal (gray) and model thereof (black) are shown at the bottom.

For every shift, a model was created using the following formula,

$$model = \sum_{i=1}^4 P_i \times \beta_i \quad (1)$$

where P represents a predictor and β the corresponding regression beta values. In this formula, the summation is over 4 instead of 3 predictors due to the addition of the intercept.

The model with the highest variance explained was used for further analysis. The variance explained was calculated by the following formula,

$$(SS_{mean} - SS_{model}) / SS_{mean} \quad (2)$$

where SS_{mean} is the sum of squared differences between the signal and the mean of that signal, and SS_{model} the sum of squared differences between the signal and the model.

The chosen model was tested for significance of explaining the data by using the analysis of variance statistic, with $\alpha=0.05$ (corrected for false discovery rate; FDR). Note that for such a statistical test the degrees of freedom in the denominator is $n-k$, where k is the number of predictors (including the intercept), and n the number of observations, assuming that all sample points are independent measurements. However, this would lead to an overestimation of the significance since the frequency conversion creates dependence among consecutive sample points over some time span. In addition, the HFB signal is a proxy for the underlying neural events, which by itself have an inherent temporal width. Therefore, n (the number of observations) was not set to one

observation per sample point but to one observation per second of signal time (n = the total signal duration in seconds). This one-second value was chosen to fully capture a neural impulse response peak (max fwhm was 0.76 seconds).

For electrodes for which the model explained the data significantly, the beta values of the three predictors, resulting from the regression analysis, were tested for significance by converting them to t-values using the following formula:

$$T_p = \frac{\beta_p}{se(\beta_p)} \quad (3)$$

where β_p is the beta value for a predictor, and $se(\beta_p)$ was calculated as follows;

$$se(\beta_p) = \sqrt{C_{jj}} \quad (4)$$

where C_{jj} is the diagonal value of the variance covariance matrix for the corresponding predictor.

The calculated t-value was subsequently converted to a one-sided p-value. The significance level was set to $\alpha = 0.05$, FDR corrected. Based on this, electrodes were assigned to one of six classes or deemed non-responsive. Electrodes that showed a significant beta value for the first predictor but not the other two predictors were classified as transient at voice onset (class 1). Electrodes that showed a significant beta value for the first and last predictor but not the second predictor were classified as transient at voice onset and voice offset (class 2) and if only the third predictor was significant, it was classified as transient at voice offset (class 3). Electrodes that had a significant positive beta value for the second predictor were classified as either only sustained (class 4), sustained with a peak at voice onset (class 5) or sustained with a peak at voice onset and voice offset (class 6), depending on whether the first and or third predictor were also significant. A sustained response with a peak only at voice offset was not found in our data and is therefore not mentioned further. If no predictors were significant, an electrode was classified as non-responsive.

For visualization purposes, the signal was epoched (see Figure 2) and trials of the same condition (1, 2 or 3 seconds) were averaged. This led to an average HFB-trace per electrode and per condition. These were subsequently averaged over electrodes and over subjects for each condition and class separately (see Figure 3).

Finally, we performed a correlation analysis between the model shift timing that explained the data best, serving as an indication of the HFB response onset timing relative to a movement (as indicated by voice onset or voice offset) and the anatomical dorsal-ventral localization. This was done to see if ventral areas showed relatively later responses than more dorsal areas as suggested by Bouchard and colleagues (2013).

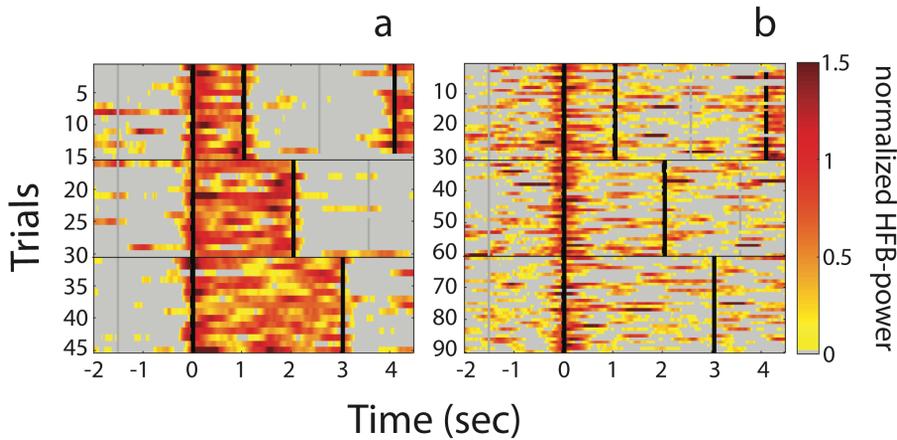


Figure 2. Two examples of the inter-trial variability. (a) Example of an electrode with a sustained response (see arrow in Figure 4b). (b) Example of an electrode with a transient response (see arrow in Figure 4c). On the x-axis, time is indicated with zero being voice onset. On the y-axis, individual 1, 2 and 3 second trials are shown. The color-scale indicates the normalized HFB-power. The vertical black lines indicate voice onset or voice offset. The gray vertical lines indicate the trial duration cue. Note that in the one-second trials, part of the subsequent trial is visible towards the end. Also, note that the trials are aligned relative to voice onset, and that deviations in voice offset from the intended duration were corrected. In (b) twice the number of trials is shown since this subject performed the task twice.

Results

Subjects performed the task well and no trials were removed based on behavioral performance. The average reaction time was 0.46 seconds ($SD=0.23$) and duration of pronunciation deviated from the intended speech duration, by -0.25 ($SD=0.33$) seconds on average, meaning that subjects usually pronounced the vowel a little shorter than instructed.

Each electrode that had a significant response (i.e. for which the model explained the data significantly) was classified as belonging to one of six classes as mentioned above (see method section E and Figure 3). Of all electrodes, on average (over subjects and weighted by number of sensorimotor cortex electrodes per subject) 29.1% (59/203, $SD=7.4\%$) were classified as having a transient response without an additional sustained response, and 17.2% (35/203, $SD=8.5\%$) as showing a sustained response with or without a transient peak. These sustained responses were related to the duration of the action, showing an increasing brain response duration with increasing vowel duration (Figure 3). The remaining electrodes (53.7%, 112/203, $SD=15.1\%$) did not show a significant response to this analysis. Of the electrodes that were classified as transient, 59.3% (35/59, $SD=13.9\%$) was classified as transient at voice onset only, 30.5% (18/59, $SD=13.0\%$) as

transient at voice onset and offset, and 10.2% (6/59, SD=5.3%) as transient at voice offset only. Interestingly, the transient electrodes that showed a peak at voice offset seem to have reduced activity between voice onset and offset compared to baseline, whereas the electrodes that only showed a peak at voice onset did not (see Figure 3a,c,e). Of all electrodes that were classified as sustained, 54.3% (19/35, SD=18.7%) also showed a peak at voice onset, 34.3% (12/35, SD=19.9%) a peak at voice onset and offset, and another 11.4% (4/35, SD=14.5%) was sustained without a clear peak of activity at voice onset or offset.

When looking at the anatomical localization of electrodes with different response profiles (Figure 4), no clear anatomical organization could be observed between transient and sustained electrodes, except for the subjects with the HD electrodes (subject C & D). For subject C, a cluster of transient responses seemed to be more anteriorly located, and a cluster of sustained responses more posteriorly. Subject D showed two clusters of sustained responses, one located ventral-posteriorly and the other more dorsal-anteriorly. For these subjects, we also plotted the timing of model shift for each sensorimotor electrode, serving as a marker for neural activity onset (see Figure 5 for the model shift timing). Notably, most electrodes started their activity before voice onset and showed a peak in activity mostly on or just after voice onset, see also Figure 3 for this. Furthermore, the more ventral electrodes seemed activate later than the dorsal electrodes for subject C, with a negative relation between the ventral-dorsal localization and response onset ($R=-0.34$, $p=0.08$), see Figure 5. For subject D, the same result was visible ($R=-0.39$, $p=0.01$) although there also seemed to be an anterior-posterior division with later responses mostly located posteriorly.

Discussion

In this study, we showed that the neuronal activity patterns recorded from the sensorimotor cortex can be directly related to action duration in some locations, whereas in other locations, during the same action, neuronal activity is transient, with a peak in HFB activity at voice onset and/or offset. To our knowledge, this is the first time that both transient and sustained neuronal dynamics of sensorimotor activity have been simultaneously and systematically mapped with respect to speech-actions of different durations in humans.

When looking at the results obtained in the three subjects with standard clinical grids over the sensorimotor mouth area, populations with sustained and transient responses seemed to be spatially scattered, without a clear anatomical organization or separation. Interestingly, however, data from the patients with a high-density grid allowed us to take a more detailed look at regional differences between response profiles and revealed that in the ventral sensorimotor mouth area, electrodes showing sustained responses occurred more posteriorly. Also, for one subject that had HD coverage more dorsally, a cluster of sustained responses was seen more anteriorly. These anatomical differences have not been shown before and might suggest that populations of neurons with transient and sustained response profiles are anatomically separated. Moreover, the high-density grid data revealed that activity that was associated with the onset of pronunciation usually started before voice onset, with a peak of activity close to voice onset for the more dorsal electrodes and usually somewhat later for the ventral electrodes. This result corresponds to the findings from Bouchard and colleagues (2013) who found that articulators that are represented more dorsally (i.e.

lips and jaw), show earlier responses, close to voice onset or before that, and features of the tongue position (located ventrally) are represented later, i.e. after voice onset. They also showed that the responses for the larynx (representation both ventrally and dorsally) were locked to voice onset, which could explain some of the early ventral responses in our data. Taken together, our data suggest that there is an underlying distinction between sustained and transient responses and also between early and late responses, with respect to their anatomical location, but more high-density ECoG data will be needed to confirm these findings.

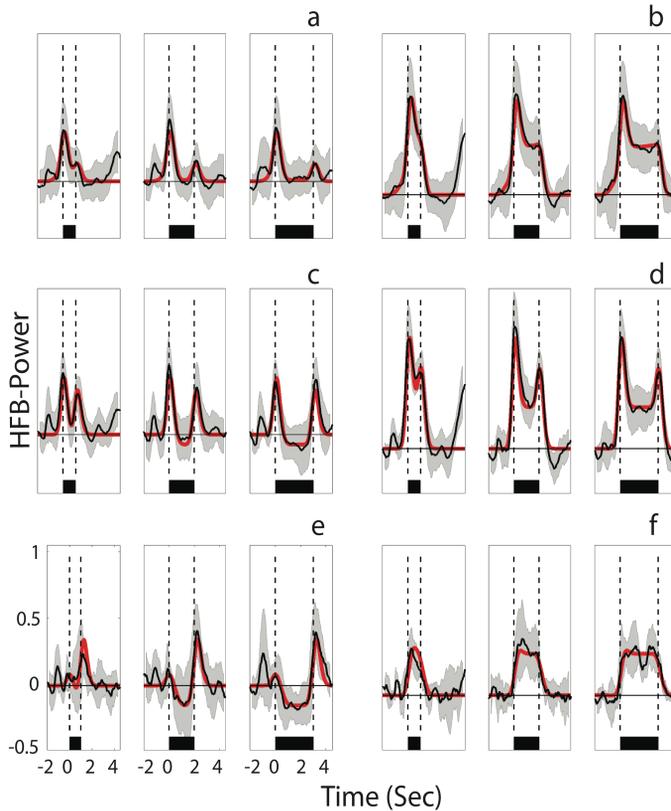


Figure 3. Response types. A visualization of the six response patterns observed in the data, averaged over all subjects. (a) Transient at voice onset, (b) sustained and transient at voice onset, (c) transient at voice onset and offset, (d) sustained and transient at voice onset and offset, (e) transient at voice offset, (f) sustained. In all panels, time is indicated on the x-axis, with zero being voice onset. On the y-axis, the normalized HFB-power is indicated. The mean of all traces that were classified as belonging to a specific class is shown in black, with the standard deviation indicated in light gray. For every electrode, based on the beta values of the three predictors a best fit model was created (see Figure 1) and the red line shows the mean of all electrode models. The horizontal black bar shows the duration of speech, being 1, 2 and 3 seconds from left to right, respectively. Vertical dashed lines indicate voice onset and voice offset. Note that on average, neural activity already started before voice onset.

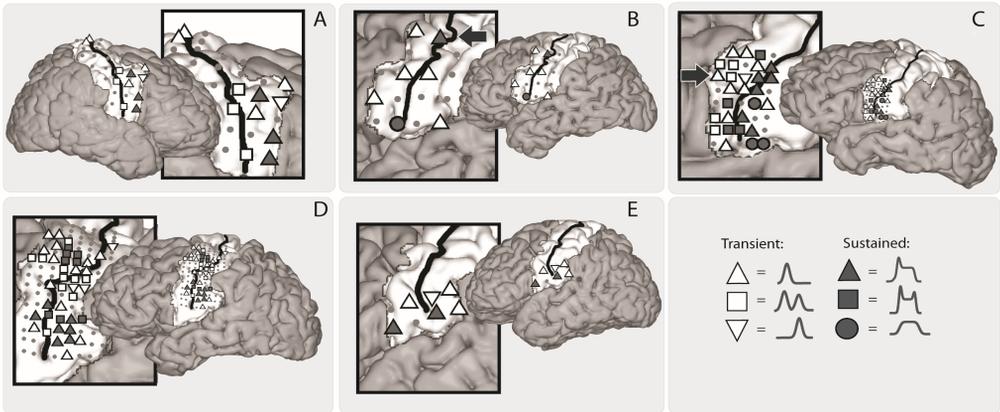


Figure 4. Response patterns. A visualization of the electrode positions (grey dots) and the response types, per subject (A-E). Symbols indicate the response type of an electrode. Arrows in B and C indicate the electrodes that are shown in Figure 2.

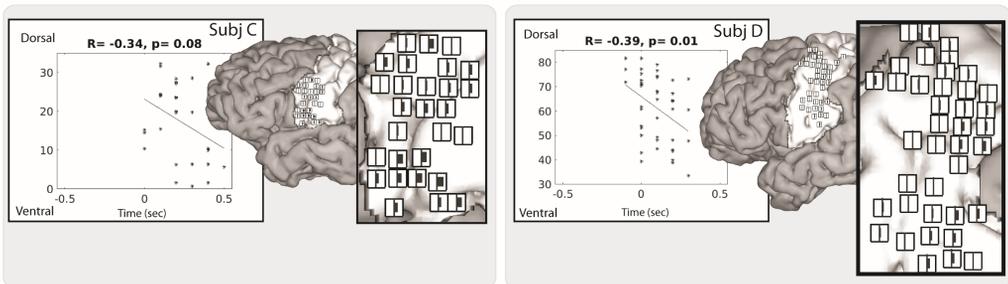


Figure 5. Response onset. A visualization of the model shift timing for the subjects with HD-grid implantation (subject C and D). Scatter plots with time in seconds on the x-axis (zero being voice onset), and the ventral-dorsal MRI coordinates on the y-axis, indicate the correlation between model shift (a marker for the response onset) and ventral-dorsal brain position for each electrode. The anatomy plots show the locations of the electrodes and their timing relative to voice onset. The vertical mid-line represents voice onset and the horizontal bars indicate the model shift (black bar to the right indicates a peak of activity after voice onset, black bar to the left indicates peak activity prior to voice onset, with a larger bar indicating a larger temporal difference with voice onset). Note that the ventral parts are not well covered in subject D while this is the area in subject C with the latest responses. Note also that these timings correspond to timing of the first peak, meaning that a shift time on or near voice onset (vertical midline) reflects a rising activation before voice onset.

Another phenomenon we observed was the occurrence, in both precentral and postcentral areas, of a (second) transient peak close to voice offset in some electrodes. In addition, for several electrodes that were classified as having only a peak at voice onset, it seemed that there was a small second peak (although not statistically significant). It could be that this voice offset peak was

sometimes present but masked due to noise. Interestingly, Ball and colleagues (2008) described similar offset-related increases in HFB power for arm movements and Hermes and colleagues (Hermes et al. 2012) found the same result for finger movements, mainly within the postcentral gyrus which is related to somatosensory functions (Penfield and Boldrey 1937), and they attributed this to the notion that some cells in this area can fire with two directions of movements (Soso and Fetz 1980). Also cells in the precentral gyrus have been found to be active with two directions; during flexion and/or extension of joints (Fetz et al. 1980). It is likely that the voice offset related peaks we observed are associated with articulators moving back to their rest position. Furthermore, we found that the electrodes that showed this offset related peak displayed a reduction in activation compared to baseline between voice onset and offset. Previous functional magnetic resonance imaging (fMRI) and transcranial magnetic stimulation (TMS) studies have linked motor related deactivation responses (for hand movements) in the ipsilateral cortex to inhibition of the opposite hand and claim this to be necessary to reduce interfering movements (Allison et al. 2000). Possibly, inhibition of neurons responsible for movements which are offset related and oppose the articulator positions during the pronunciation cause a deactivation in the HFB power.

Correspondence to Previous Research

We showed that different patterns of activity can be recorded with electrodes placed on the surface of the brain and our results are in correspondence to earlier primate studies. For instance, we found that both transient-only and transient-followed-by-sustained responses show a peak of activity at movement onset, followed by a decrease in activity (which was larger in transient responses than in sustained responses). This is in correspondence to earlier findings of single cell recordings in primates that first show a burst of activity during a dynamic movement phase followed by a decrease to a steady level of elevated activity during a tonic phase. In contrast, other cells are highly active only during the tonic phase after a movement (Cheney and Fetz 1980), which corresponds to the sustained only responses that we found.

Previous work (Crone et al. 1998; Conant et al. 2018) suggested that both duration related and duration unrelated responses are involved in hand and speech movements in humans, which is in correspondence to our data. We extend those findings, in that we experimentally show that the sustained responses are related to the duration of an action and that transient responses are related to movement onset or offset. Furthermore, we mapped these two types of responses on the sensorimotor cortex and show their anatomical positioning. However, in contrast to the findings of Crone and colleagues (1998), who indicated that transient responses are more associated with higher-gamma frequencies (75-100 Hz) and sustained responses with low-gamma frequencies (40-50 Hz), we observed sustained responses in the higher frequencies. Although they suggest that lower frequencies are more involved in, for instance, motor output or sustained attention and higher frequencies in motor planning or initiation of movements, our results suggest that high frequency power can be associated with motor output as well. Task differences could, however, explain this inconsistency (see below for discussion).

Function of Sustained and Transient Responses

One possible explanation for the sustained responses we observed is that they reflect the movement of body parts that are continuously moving or that apply continuous force during speech (in contrast to transient responses for non-continuously moving body parts). For instance, pulmonic egressive airstreams (air pushed from the lungs) enable speech (Rietveld and van Heuven 2001) and during continuous speech, lung volume decreases steadily, which is realized mostly by internal and external intercostal muscle activity (Draper et al. 1959). This might explain why in the study by Crone and colleagues (1998), sustained responses were not found for higher-gamma frequencies, since subjects did not constantly move. They were asked to make a movement as fast as possible (e.g. tongue protrusion) and keep it in that position for a couple of seconds. Alternatively, sustained responses may reflect continuous somatosensory feedback associated with sustained sound production. The data from our subjects with HD-grids may support this possibility in that clusters of sustained responses were observed in more posterior positions (towards the somatosensory cortex) in the ventral parts of the sensorimotor cortex. Indeed, the postcentral gyrus is mainly involved in controlling somatosensory processing, whereas the precentral gyrus is more involved in motor execution (Penfield and Boldrey 1937). However, there is also overlap in function between pre- and postcentral gyrus activity with respect to both motor and sensory responses (Fetz et al. 1980; Soso and Fetz 1980; Penfield and Boldrey 1937). Therefore, some deviations from this classical pre- and postcentral division for sustained (and transient) responses can be expected. In both subjects with HD-coverage, sustained responses in the more ventral parts of the sensorimotor cortex were predominantly observed posterior to the central sulcus, suggesting that for ventral areas there might be an anterior-posterior distinction for sustained and transient responses (although more data is needed to investigate this). However, in one of these subjects, in the more dorsal areas, sustained responses were also located anterior to the central sulcus. Although this contradicts the anterior-posterior division from the ventral parts, we only observed this for one subject. It is therefore unclear if this finding was subject specific or a general phenomenon. A third possibility for the role of sustained responses could be the encoding of the position of a body part. Bouchard and colleagues (2016) have suggested that the HFB-power is correlated with the position of the lips, which could suggest that the sustained responses reflect articulator position. However, since their results are correlational, no causal conclusion can be derived from this. Support for the position hypothesis, however, is also provided by single cell recordings in both primates and humans, that show position specific activity for sensorimotor cortex cells (Truccolo et al. 2008; Wang et al. 2007) especially during a static phase after a movement (Cheney and Fetz 1980; Fetz et al. 1980; Soso and Fetz 1980).

Transient responses, on the other hand, can be expected if neuronal activity is related to initiation of the movement, and could therefore be associated with articulators or body parts that move only during the initiation or ending of sound production (in contrast to sustained responses for continuously moving body parts). For instance, the tongue has been shown to start moving just before voice onset, stay in a steady position throughout single syllable pronunciation and then move back just after voice offset (Bouchard et al. 2016). Another explanation would be that transient responses reflect movement planning. Indeed, non-human primate vocalization research (Coudé et al. 2011; Hage and Nieder 2013) as well as research in songbirds (Tang et al. 2014) has described

responses often preceding or surrounding voice onset. These profiles were most often seen in premotor areas, an area generally associated with motor planning (Kandel et al. 2000; Culham 2015; Rizzolatti and Luppino 2015; Vargas-Irwin et al. 2015). In our study, we found transient responses in the precentral and postcentral parts of the sensorimotor cortex, with activity usually starting before, but peak activity just on or after, voice onset (and/or voice offset). In correspondence to our results, activity of some neurons in the precentral gyrus have been found to depend on movement instructions during a preparatory (non-moving) phase before monkeys get a cue to move (push versus pull an object with their arm). Errors in movements can be predicted from these neurons before the actual movement (Tanji and Evarts 1976). This indicates that indeed, besides in the premotor areas, motor planning may be represented within the precentral gyrus as well.

Implications for Brain-Computer Interfaces

Our findings may help in the development of brain-computer interfaces (BCIs) for subjects that suffer from complete or nearly complete paralysis. Although it has been shown that speech units can be classified from sensorimotor cortex activity (Kellis et al. 2010; Mugler et al. 2014; Brumberg et al. 2011; Herff et al. 2015), accuracy scores usually do not meet the standards for home use BCI-controlled language communication applications. This might potentially improve when more is known about the relationship between small speech units, and variations in pronunciation of these, and the brain signal. Our current study contributes to our understanding of the sensorimotor cortex and the relation between neuronal activity and behavioral output, which could lead to better classification of syllables, phonemes and eventually words and sentences.

Limitations

There are several limitations to this study that need to be addressed. First, only a limited number of subjects participated in this study and only two subjects had high-density electrode coverage. As described before, only subject C & D (with HD-grids) showed separate clusters of transient and sustained responses. It could be that, due to the sparse sampling with clinical grids, we missed active cortical sites in the other three patients. Alternatively, one could argue that in two of the three subjects with the clinical grids, statistical power was low, given that they performed only one run (and therefore half of the trials). However, this does not explain why the one subject with clinical grids who performed two runs, also did not show a functional organization. Also, since the electrodes in the clinical grids are larger, they record from a larger population of neurons than the HD-grid electrodes, which may lead to spatial blurring, mixing of response profiles and a smaller signal to noise ratio (for a simulation of this see Camuñas-Mesa and Quiroga 2013). It may be speculated that, because of the larger number of electrodes per square cm and the fact that these electrodes are also smaller, high-density recordings are especially suitable to accurately study the detailed organization of the sensorimotor cortex.

Second, we corrected the data for incorrect response timing (by interpolation and down sampling) to be able to compare the responses of different trials and different subjects. We don't believe this procedure has induced the differences between sustained and transient responses we

observed, since during individual trials, both sustained and transient responses were found (in different locations) while the correction was the same for all electrodes.

Third, since the subjects heard their own voice during the task, we cannot rule out that the auditory stimulation contributed to the cortical responses. However, since sustained movement related responses in the sensorimotor cortex during single cell recordings in primates have been found in the absence of auditory stimulation we argue that the current results cannot be solely attributed to auditory stimulation.

Conclusion

We demonstrate here that some focal regions of the sensorimotor mouth area show sustained responses associated with action duration, whereas other sites show transient responses coupled to movement onset and/or offset. Sustained responses may be associated with continuous movement or force, somatosensory feedback or articulator position, whereas transient responses could reflect the (initiation of) short-duration movements or motor planning. We believe our findings warrant further research into the nature of cortical responses during elementary articulations, which may improve our understanding of the cortical representation of complex motor actions such as speech, and thereby improve decodability for brain-computer interfaces.

Acknowledgements

The authors thank the staff of the clinical neurophysiology department, the neurosurgeons and the subjects for their contribution.

References

- Allison JD, Meador KJ, Loring DW, et al (2000) Functional MRI cerebral activation and deactivation during finger movement. *Neurology* 54:135–135. doi: 10.1212/WNL.54.1.135
- Ball T, Demandt E, Mutschler I, et al (2008) Movement related activity in the high gamma range of the human EEG. *NeuroImage* 41:302–310. doi: 10.1016/j.neuroimage.2008.02.032
- Bleichner MG, Jansma JM, Salari E, et al (2015) Classification of mouth movements using 7 T fMRI. *J Neural Eng* 12:066026. doi: 10.1088/1741-2560/12/6/066026
- Boersma P (2002) Praat, a system for doing phonetics by computer. *Glott Int* 5.
- Booij G (1999) *The Phonology of Dutch*. Clarendon Press
- Bouchard KE, Conant DF, Anumanchipalli GK, et al (2016) High-Resolution, Non-Invasive Imaging of Upper Vocal Tract Articulators Compatible with Human Brain Recordings. *PLOS ONE* 11:e0151327. doi: 10.1371/journal.pone.0151327
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–332. doi: 10.1038/nature11911
- Branco M, Gaglianese A, Hermes D, et al (2016) Pipeline for ECoG electrode localization on brain surface: towards a one click approach. In: *Proceedings of the 6th International Brain-Computer Interface Meeting*, pp. 154. doi: 10.3217/978-3-85125-467-9-154
- Branco MP, Freudenburg ZV, Aarnoutse EJ, et al (2018) Optimization of sampling rate and smoothing improves classification of high frequency power in electrocorticographic brain signals. *Biomed Phys Eng Express* 4:045012. doi: 10.1088/2057-1976/aac3ac
- Brumberg JS, Wright EJ, Andreasen DS, et al (2011) Classification of Intended Phoneme Production from Chronic Intracortical Microelectrode Recordings in Speech-Motor Cortex. *Front Neurosci* 5. doi: 10.3389/fnins.2011.00065
- Bruns A (2004) Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? *J Neurosci Methods* 137:321–332. doi: 10.1016/j.jneumeth.2004.03.002
- Camuñas-Mesa LA, Quiroga RQ (2013) A Detailed and Fast Model of Extracellular Recordings. *Neural Comput* 25:1191–1212. doi: 10.1162/NECO_a_00433
- Cheney PD, Fetz EE (1980) Functional classes of primate corticomotoneuronal cells and their relation to active force. *J Neurophysiol* 44:773–791
- Conant DF, Bouchard KE, Leonard MK, Chang EF (2018) Human sensorimotor cortex control of directly-measured vocal tract movements during vowel production. *J Neurosci* 2382–17. doi: 10.1523/JNEUROSCI.2382-17.2018
- Coudé G, Ferrari PF, Rodà F, et al (2011) Neurons Controlling Voluntary Vocalization in the Macaque Ventral Premotor Cortex. *PLOS ONE* 6:e26822. doi: 10.1371/journal.pone.0026822
- Crone NE, Hao L, Hart J, et al (2001) Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology* 57:2045–2053. doi: 10.1212/WNL.57.11.2045
- Crone NE, Miglioretti DL, Gordon B, Lesser RP (1998) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* 121:2301–2315. doi: 10.1093/brain/121.12.2301
- Culham J (2015) Cortical Areas Engaged in Movement: Neuroimaging Methods - In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Elsevier, Oxford, pp 21–29
- Dechent P, Frahm J (2003) Functional somatotopy of finger representations in human primary motor cortex. *Hum Brain Mapp* 18:272–283. doi: 10.1002/hbm.10084
- Donoghue JP, Sanes JN, Hatsopoulos NG, Gaál G (1998) Neural Discharge and Local Field Potential Oscillations in Primate Motor Cortex During Voluntary Movements. *J Neurophysiol* 79:159–173
- Draper MH, Ladefoged P, Whitteridge D (1959) Respiratory Muscles in Speech. *J Speech Lang Hear Res* 2:16–27. doi: 10.1044/jshr.0201.16
- Espadaler J, Rogić M, Deletis V, et al (2012) Representation of cricothyroid muscles at the primary motor cortex (M1) in healthy subjects, mapped by navigated transcranial magnetic stimulation (nTMS). *Clin Neurophysiol* 123:2205–2211. doi: 10.1016/j.clinph.2012.04.008
- Fetz EE, Finocchio DV, Baker MA, Soso MJ (1980) Sensory and motor responses of precentral cortex cells during comparable passive and active joint movements. *J Neurophysiol* 43:1070–1089
- Fischl B (2012) FreeSurfer. *NeuroImage* 62:774. doi: 10.1016/j.neuroimage.2012.01.021
- Georgopoulos, Ashe J, Smyrnis N, Taira M (1992) The Motor Cortex and the Coding of Force. *Science* 256:1692–1695
- Georgopoulos, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J Neurosci* 2:1527–1537

- Hadoush H, Sunagawa T, Nakanishi K, et al (2011) Motor somatotopy of extensor indicis proprius and extensor pollicis longus. *Neuroreport* 22:559–564. doi: 10.1097/WNR.0b013e328348e750
- Hage SR, Nieder A (2013) Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat Commun* 4:2409. doi: 10.1038/ncomms3409
- Herff C, Heger D, de Pestors A, et al (2015) Brain-to-text: decoding spoken phrases from phone representations in the brain. *Neural Technol* 9:217. doi: 10.3389/fnins.2015.00217
- Hermes D, Miller KJ, Noordmans HJ, et al (2010) Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J Neurosci Methods* 185:293–298. doi: 10.1016/j.jneumeth.2009.10.005
- Hermes D, Siero JCW, Aarnoutse EJ, et al (2012) Dissociation between Neuronal Activity in Sensorimotor Cortex and Hand Movement Revealed as a Function of Movement Rate. *J Neurosci* 32:9736–9744. doi: 10.1523/JNEUROSCI.0357-12.2012
- Hermes DJ (1990) Vowel-onset detection. *J Acoust Soc Am* 87:866–873. doi: 10.1121/1.398896
- Kandel E, Schwartz J, Jessell T (2000) *Principles of Neural Science*. McGraw-Hill Medical
- Kellis S, Miller K, Thomson K, et al (2010) Decoding spoken words using local field potentials recorded from the cortical surface. *J Neural Eng* 7:056007. doi: 10.1088/1741-2560/7/5/056007
- Liu Y, Coon WG, Pestors A de, et al (2015) The effects of spatial filtering and artifacts on electrocorticographic signals. *J Neural Eng* 12:056008. doi: 10.1088/1741-2560/12/5/056008
- Manning JR, Jacobs J, Fried I, Kahana MJ (2009) Broadband shifts in LFP power spectra are correlated with single-neuron spiking in humans. *J Neurosci Off J Soc Neurosci* 29:13613. doi: 10.1523/JNEUROSCI.2041-09.2009
- Miller KJ, Leuthardt EC, Schalk G, et al (2007) Spectral Changes in Cortical Surface Potentials during Motor Movement. *J Neurosci* 27:2424–2432. doi: 10.1523/JNEUROSCI.3886-06.2007
- Miller KJ, Sorensen LB, Ojemann JG, Nijs M den (2009a) Power-Law Scaling in the Brain Surface Electric Potential. *PLoS Comput Biol* 5:e1000609. doi: 10.1371/journal.pcbi.1000609
- Miller KJ, Zanos S, Fetz EE, et al (2009b) Decoupling the Cortical Power Spectrum Reveals Real-Time Representation of Individual Finger Movements in Humans. *J Neurosci* 29:3132–3137. doi: 10.1523/JNEUROSCI.5506-08.2009
- Moran DW, Schwartz AB (1999) Motor Cortical Representation of Speed and Direction During Reaching. *J Neurophysiol* 82:2676–2692
- Mugler EM, Patton JL, Flint RD, et al (2014) Direct classification of all American English phonemes using signals from functional speech motor cortex. *J Neural Eng* 11:035015. doi: 10.1088/1741-2560/11/3/035015
- Pei X, Leuthardt EC, Gaona CM, et al (2011) Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *NeuroImage* 54:2960–2972. doi: 10.1016/j.neuroimage.2010.10.029
- Penfield W, Boldrey E (1937) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain J Neurol* 60:389–443. doi: 10.1093/brain/60.4.389
- Ray S, Maunsell JHR (2011) Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLoS Biol* 9:e1000610. doi: 10.1371/journal.pbio.1000610
- Rietveld A, van Heuven V (2001) *Algemene Fonetiek*, 2nd edn. Coutinho, Bussum
- Rizzolatti G, Luppino G (2015) Premotor Cortex - In: *International Encyclopedia of the Social & Behavioral Sciences* (Second Edition). Elsevier, Oxford, pp 846–851
- Siero JC, Hermes D, Hoogduin H, et al (2014) BOLD matches neuronal activity at the mm scale: A combined 7 T fMRI and ECoG study in human sensorimotor cortex. *NeuroImage* 101:177–184. doi: 10.1016/j.neuroimage.2014.07.002
- Soso MJ, Fetz EE (1980) Responses of identified cells in postcentral cortex of awake monkeys during comparable active and passive joint movements. *J Neurophysiol* 43:1090–1110
- Tang C, Chehayeb D, Srivastava K, et al (2014) Millisecond-Scale Motor Encoding in a Cortical Vocal Area. *PLoS Biol* 12:e1002018. doi: 10.1371/journal.pbio.1002018
- Tanji J, Evarts EV (1976) Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. *J Neurophysiol* 39:1062–1068
- Taylor CSR, Gross CG (2003) Twitches Versus Movements: A Story of Motor Cortex. *The Neuroscientist* 9:332–342. doi: 10.1177/1073858403257037
- Towle VL, Yoon H-A, Castelle M, et al (2008) ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain* 131:2013–2027. doi: 10.1093/brain/awn147
- Truccolo W, Friehs GM, Donoghue JP, Hochberg LR (2008) Primary Motor Cortex Tuning to Intended Movement Kinematics in Humans with Tetraplegia. *J Neurosci* 28:1163–1178. doi: 10.1523/JNEUROSCI.4415-07.2008
- Vargas-Irwin CE, Franquemont L, Black MJ, Donoghue JP (2015) Linking Objects to Actions: Encoding of Target Object and Grasping Strategy in Primate Ventral Premotor Cortex. *J Neurosci* 35:10888–10897. doi: 10.1523/JNEUROSCI.1574-15.2015
- Wang PT, King CE, McCrimmon CM, et al (2014) Electrocorticogram encoding of upper extremity movement

- duration. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp 1243–1246
- Wang PT, McCrimmon CM, King CE, et al (2017) Characterization of electrocorticogram high-gamma signal in response to varying upper extremity movement velocity. *Brain Struct Funct* 222:3705–3748. doi: 10.1007/s00429-017-1429-8
- Wang W, Chan SS, Heldman DA, Moran DW (2007) Motor Cortical Representation of Position and Velocity During Reaching. *J Neurophysiol* 97:4258–4270. doi: 10.1152/jn.01180.2006





Chapter 2

*“The songs are in your eyes, I see them when you smile.
I’ve had enough, I’m not giving up on a miracle drug.”*

Chapter 2

Repeated Vowel Production Affects Features Of Neural Activity In Sensorimotor Cortex

E. Salari, Z.V. Freudenburg, M.J. Vansteensel, N.F. Ramsey

Published Article: Brain Topography, September 2018

Abstract

Background: The sensorimotor cortex is responsible for the generation of movements and interest in the ability to use this area for decoding speech by brain-computer interfaces has increased recently. Speech decoding is challenging however, since the relationship between neural activity and motor actions is not completely understood. Non-linearity between neural activity and movement has been found for instance for simple finger movements. Despite equal motor output, neural activity amplitudes are affected by preceding movements and the time between movements. It is unknown if neural activity is also affected by preceding motor actions during speech. **Methods:** We addressed this issue, using electrocorticographic high frequency band (HFB; 75-135 Hz) power changes in the sensorimotor cortex during discrete vowel generation. Three subjects with temporarily implanted electrode grids produced the /i/ vowel at repetition rates of 1, 1.33 and 1.66 Hz. For every repetition, the HFB power amplitude was determined. **Results:** During the first utterance, most electrodes showed a large HFB power peak, which decreased for subsequent utterances. This result could not be explained by differences in performance. With increasing duration between utterances, more electrodes showed an equal response to all repetitions, suggesting that the duration between vowel productions influences the effect of previous productions on sensorimotor cortex activity. **Discussion:** Our findings correspond with previous studies for finger movements and bear relevance for the development of brain-computer interfaces that employ speech decoding based on brain signals, in that past utterances will need to be taken into account for these systems to work accurately.

Keywords; Sensorimotor Cortex, Movement, Repetition, ECoG, Speech, Brain-Computer Interface

Introduction

The execution of everyday voluntary body movements generally occurs without effort and is the result of the concerted action of different neural processes and brain areas. The sensorimotor cortex is known to play a central role in the different aspects of the generation of movement, such as the control of body part positions, the velocity and direction of movements, applied force and the planning of motor actions (Tanji and Evarts 1976; Georgopoulos et al. 1982, 1992; Donoghue et al. 1998; Moran and Schwartz 1999; Wang et al. 2007; Truccolo et al. 2008). However, for subjects suffering from severe forms of paralysis, even the most common forms of movements, such as those involved in speech and communication can sometimes be completely absent (American Congress of Rehabilitation Medicine 1995; Smith and Delargy 2005; Posner et al. 2007). To restore communication in these subjects, brain-computer-interface (BCI) systems are being developed (Wolpaw et al. 2002). These systems may convert neural activity into written or spoken computer output, and sensorimotor cortex activity related to speech has been shown useful in an attempt to identify, from the neural signals, which sound or word a user may want to communicate (Kellis et al. 2010; Mugler et al. 2014; Herff et al. 2015; Ramsey et al. 2018). These attempts usually rely on the assumption that each specific sound or word is associated with a unique neural signature. Imaging and patient studies, however, have shown that repeating a movement in a discrete way (with short pauses between each movement) may involve different brain areas than performing the same movements in a continuous way (without short pauses between each movement; Kennerley et al. 2002; Spencer et al. 2003; Miall and Ivry 2004; Schaal et al. 2004), even though the movements are almost identical. Moreover, there is evidence for a non-linear relationship between movement-performance and neural activity in the sensorimotor cortex. Various studies have suggested that previous actions influence the neural activity associated with subsequent actions, if spaced close enough together (Miezin et al. 2000; Soltysik et al. 2004). Indeed, during repeated finger movements, the amplitude of sensorimotor neural activity, as measured with fMRI and electrocorticography (ECoG), was shown to decline over repetitions, despite equal movement output (Hermes et al. 2012b; Siero et al. 2013; for a comparison between BOLD and ECoG see: Logothetis et al. 2001; Hermes et al. 2012a; Siero et al. 2014).

Importantly, the studies mentioned above focused on hand and finger movements and it remains to be determined whether the observed complex and non-linear relationship between movement and underlying neural activity is a general feature of the sensorimotor cortex, or whether it is specific to the areas involved in hand movement. Especially relevant in this respect is our previous finding that different parts of the sensorimotor cortex show different response profiles to the same *speech* movement. Some cortical foci show sustained neural activity during a sustained motor speech action whereas in other locations responses are transient during the same movement (Salari et al. 2018). This finding indicates that the relationship between neural activity and overt speech behaviour differs between subareas of the sensorimotor cortex. It could be speculated that the presence, or absence, of a non-linear relationship between neural responses and behavioural output during repeated movements is specific for cortical foci as well.

With the current study, we aimed to obtain a better understanding of the link between speech pronunciation and underlying sensorimotor cortex activity. This is of interest for BCIs that employ

neural signal changes related to (attempted) speech. If the neural signal associated with a specific (attempted) pronunciation would be affected by previous speech actions, the same word or sound may be related to a diversity of neural signatures, which have to be taken into account for a sensorimotor-speech-BCIs to function accurately.

In this study, we investigated the relationship between repeated orofacial movements during speech, and sensorimotor brain activity. We recorded neural signals in three subjects while they pronounced the same vowel multiple times, at different repetition rates. Neural activity was recorded with subdural ECoG electrodes, which allows for recording at high temporal resolution and with high spatial specificity (Siero et al. 2014). We focused on frequencies in the range of 75-135 Hz, which are known to have a spatially specific relationship with (speech and articulator) movements (Crone et al. 1998; Miller et al. 2007; Bouchard et al. 2013), and which are thought to reflect underlying neural population firing (Manning et al. 2009; Miller et al. 2009; Ray and Maunsell 2011). We focused mostly on the ventral parts of the sensorimotor cortex as this area has previously been shown to be responsible for the generation of speech movements (Penfield and Boldrey 1937; Crone et al. 2001; Towle et al. 2008; Pei et al. 2011; Bouchard et al. 2013) and has been the focus of BCI-studies for the classification of speech sounds (see for instance Kellis et al. 2010; Mugler et al. 2014; Herff et al. 2015; Ramsey et al. 2018) and articulator movements (Bleichner et al. 2015).

Methods

Participants

Subjects included in this study (n=3, 2 females, 19, 41 and 30 years old respectively) were implanted with subdural clinical ECoG electrodes for epilepsy treatment at the University Medical Center Utrecht. All subjects had an additional high-density (HD) electrode grid placed over the sensorimotor cortex (SMC; left for subject A & B and right for subject C). These grids were exclusively placed for research purposes with the subject's consent, over an area that was not clinically relevant. For subject A & B, the inter-electrode distance of the HD grid was 4 mm with an exposed electrode diameter of 1 mm for subject A and 1.17 mm for subject B. For subject C, the inter-electrode distance was 3 mm with an exposed electrode diameter of 1 mm. Only the HD electrodes were used for the current analysis.

This research was approved by the ethics committee of the University Medical Center Utrecht. All participants gave written informed consent in accordance with the Declaration of Helsinki (2013).

Task

Participants were asked to produce the /i/ vowel repeatedly at different rates (see below), guided by instructions that were visually presented on a computer screen that was placed at a distance of approximately 1 m from the participant. A trial started with an indication of the production speed by a visual cue. Subsequently, to guide the participants in producing the sound at the correct speed, the letters 'ie', corresponding in Dutch to the /i/ sound, were repeatedly visually presented for 300 ms at a rate of 5, 4, or 3 times in 3 seconds (1.66, 1.33 and 1 Hz). These repetition rates were chosen

as they were relatively easy to perform (not too slow or too fast) and because previous research for finger movements has shown that repetition effects are mostly apparent at rates of 1Hz or higher (Hermes et al. 2012b). During the inter-trial interval (1800 ms), a fixation cross was presented. Trials of different rates were randomized and each rate was repeated 26 times, divided over two recording sessions. Any trial for which the number of pronunciations was incorrect, was excluded from the analyses.

Data Acquisition & Preprocessing

Brain data was recorded and preprocessed as described previously (Salari et al. 2018). In short, ECoG data was recorded (number of electrodes: 64 for subject A and 128 for subject B & C) at a sampling frequency of 512 Hz, 2048 Hz (subject A; Micromed, Treviso, Italy), or 2000Hz (subject B & C; Blackrock Microsystems LLC, Salt Lake City, USA). Different sampling frequencies were used due to the availability of different clinical and research recording setups and the possibility, or not, to choose the most optimal sampling frequency for the current study. For subject A, the data obtained at the highest sampling frequency was down sampled such that the sampling frequencies of all datasets of subject A were the same. Electrodes in the region of interest (sensorimotor cortex) were identified by visual inspection of the electrode positions (as determined by using a post-implantation CT scan) plotted over a 3D rendering of a presurgical MRI scan (Hermes et al. 2010; Branco et al. 2018b). Sensorimotor cortex electrodes with noisy or flat signal were removed from further analysis. For the remaining electrodes, line noise (50Hz) and harmonics thereof were removed and common average re-referencing was applied. Audio recordings of the subject's pronunciation were made during the task, to identify the voice onsets and offsets and to be able to correct for possible differences in behavioral performance (see below). Voice onset and offset were determined for each vowel pronunciation, as described previously (Salari et al. 2018). Shortly, these time points were first automatically determined using a vowel detection algorithm (Hermes 1990), which was adjusted by Hermes to also detect vowel offsets. Subsequently, we corrected the on- & offsets if necessary (due to background noise for instance) using Praat software (Boersma 2002).

Matlab software (The Mathworks, Inc., Natick, MA, USA) was used for data analysis, unless specified otherwise. For all sensorimotor electrodes, the high frequency band (75-135 Hz) power was computed per sample point using a Gabor wavelet (Bruns 2004) for all frequencies between 75 and 135 Hz in bins of 1 Hz with a full width half maximum (fwhm) of 4 wavelets per frequency. Subsequently, a log transformation ($10 \cdot \log_{10}$) was applied and these results were then averaged (over frequencies) to create the HFB power signal. These signals were normalized and subsequently smoothed with a moving average window (centered around the sample point) of 0.1 seconds. This smoothing setting has been shown to be within the optimal range for accurate classification of phonemes (Branco et al. 2018a), and we used it to preserve the individual peaks per repetition in the data while reducing noise. The data from the two runs were concatenated.

Analysis of ECoG data was conducted in two steps. First, electrodes were identified and selected for further analysis based on their response to the task. Then signals from these electrodes were interrogated for vowel repetition effects.

Electrode Selection

For each electrode, we determined whether it was responsive to the task. To that purpose, we modeled the neural signal by performing a regression analysis on the whole time series. Five predictors were used in this study, each representing a transient response to one of the possible repetition numbers (max 5). Predictor 1 represents the response to all first pronunciations, the second predictor to all second pronunciations etcetera. The fourth and fifth predictors had only (predicted) responses during the trials in which there actually was a fourth and/or fifth pronunciation (Figure 1). The predictors were created by convolving a Gaussian function with an impulse function that indicated when a vowel was spoken. The full width at half maximum of the Gaussians were determined for each subject separately, as follows. First, for each electrode that had, in the trials of the slowest repetition rate, a maximum peak response higher than 1 standard deviation above the mean of the signal, we estimated the fwhm of that peak. The mean fwhm over electrodes was then used as the fwhm for the Gaussian peak of the model for the slowest production rate. For the two faster production rates, this value was adjusted to match those repetition rates by dividing it by the repetition frequency. We used the slowest repetition rate for the fwhm estimation under the assumption that this is least ‘contaminated’ with activity of other utterances. The fwhm values of the three subjects were, respectively, 0.59, 0.51 and 0.65 s for the 1 Hz repetition rate (leading to a 0.35, 0.30 and 0.39 s fwhm for the 1.66 Hz repetition rate and 0.44, 0.38 and 0.49 s fwhm for the 1.33 Hz repetition rate). Visual inspection showed that using these Gaussian widths, the neural activity could be accurately modeled for all subjects (Figure 2). Subsequently, since it is known that the HFB response onset of different areas in the brain can occur at different time-points relative to the overt motor action (Crone et al. 1998; Coudé et al. 2011; Hermes et al. 2012b; Bouchard et al. 2013), we shifted the timing of the model peaks and repeated the regression until an optimal fit was found with the data. Timing shifts ranged from 0.5 seconds before voice onset time to 0.5 seconds after voice onset, in 0.1 second increments. An electrode was considered significantly responsive to the task if it was significantly explained by the best fitting model and if the average response over trials was an *increase* in power associated with at least the first vowel production (for all three repetition rates). For subject A-C, a total number of 14, 28 and 59 electrodes were significantly active, respectively. All other electrodes were considered not-responsive (NR). Statistical analysis was conducted using analysis of variance (ANOVA; $\alpha=0.05$, false discovery rate corrected), similar as in Salari et al. (2018). For this analysis, normally, each sample point is assumed to be independent and is used as a degree of freedom. The degrees of freedom (DF) value relates to the number of independent observations but since consecutive sample points are not independent (due to the Gabor wavelet power conversion) we counted every 0.5 seconds of data as an independent sample point to not overestimate the degrees of freedom. Note that, even though the data do not necessarily meet all the assumptions for parametric testing (as discussed above), inspection of the data showed that the current analysis was useful for selecting task related electrodes.

Repetition Effects

Only the significantly responsive electrodes were used for further analysis. For these electrodes, we determined the HFB power peak amplitude for every vowel production. Since the HFB response

peak timing, with respect to voice onset, could be different for each electrode we used the shift of the model that explained the data best as the determinant for timing of the HFB response peak with respect to voice onset. Each peak amplitude was determined by taking the median of the HFB signal in a window of 0.1 sec before and 0.1 seconds after the determined peak timing. We used this value instead of the maximum value, to prevent possible noise peaks in the data to affect the results.

Correction for Performance Differences

To investigate if the duration of vowel production was influenced by the repetition rate, we performed an ANOVA for each subject with pronunciation duration (derived from the audio signal) as dependent variable and production rate as independent variable. Furthermore, we corrected for possible differences in HFB response peak amplitudes that might be caused by differences in pronunciation between repetitions. We derived four behavioral performance measures, namely (1) sound intensity, (2) lip aperture, (3) lip movement and (4) lip velocity, for this correction. Sound intensity was calculated by taking the envelope of the normalized audio signal that was recorded during the task, using the absolute value of the Hilbert transform. The sound intensity was normalized per run to make measures from different sessions comparable. Normalization (of each run) was based on the mean and standard deviation of a silent part of that run. This envelope was then smoothed with a moving window of 0.05 seconds and down sampled to 600 Hz. Lip aperture was measured by analyzing video footage of the subjects while they performed the task. For each repetition, the mean distance (in pixels) between the lips was calculated for the video frames

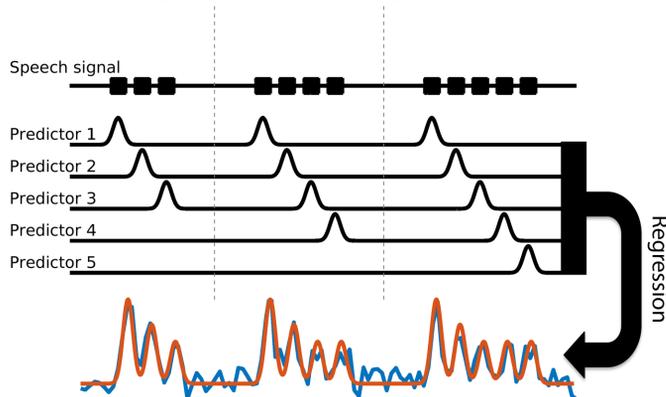


Figure 1. Model creation for electrode selection and deriving peak timings. A visual representation of the method to create from the signal (shown in blue) the model (shown in red) that was used to select significant electrodes and to determine the peak timings. Each predictor represents a transient response to one of the possible vowel production repetitions and are created by convolving a Gaussian function to an impulse at voice onset. The model was created by a regression of the five predictors to the signal. The model was used to find electrodes with a significant response to the task and to find the timing of (potential) peaks. The latter was done by shifting the predictors and repeating the regression until the best fit was found with the data.

corresponding the pronunciation. Lip movements were calculated in a similar way but the video frames during silent parts just before each pronunciation were now used. The difference between the lip aperture during silence and the upcoming pronunciation served as a measure of lip movement. Lip velocity was calculated by taking the derivative of the lip positions during the silent part before each vowel production and subsequently taking the maximum value thereof. The lip position for each analyzed frame was normalized, per run. This was done by subtracting from each lip position sample, the mean number of pixels between the lips (over analyzed frames) and dividing this by the standard deviation (of pixels over time). To see if any of the measures could explain possible differences in the brain signals, we calculated the correlation value of each of these measures with the HFB response peak amplitudes for all included electrodes. Furthermore, a principal component analysis (PCA) was performed on these measures to dissect covariance among the different measures. The principal components were used as predictors of the HFB response peak amplitudes in a regression analysis, per electrode, the result of which was subtracted from the actual HFB response peak amplitudes to regress out any performance effects on the brain data. Outliers in the HFB response peak amplitudes were disregarded and outliers in the PCA values were replaced by the average value of that component. Outliers were determined by using the 'isoutlier' function from Matlab. See supplementary Figure S1 for an indication of the variance in brain data and behavioral measures and their relation before and after correction.

Since we could not measure the tongue position in the patient subjects, we did not correct for possible differences therein. However, after the current study we repeated the task with five healthy volunteers (who signed informed consent, median age: 26 years, range 22-31 years, 1 female) and recorded their tongue position using ultrasound measures. A total of 114 echo pulse scan lines were recorded at 60.11 frames per second at a depth of 90 mm with an EchoBlaster 128 ultrasound machine. The probe was stabilized using an ultrasound headset (Articulate Instruments Ltd., 2008). The data were analyzed with Articulate Assistant Advanced software (Articulate Instruments Ltd., 2012). We then evaluated whether repeated vowel production caused systematic changes in tongue movements.

After correction for performance, the peak amplitudes of all included electrodes were averaged and grouped by repetition number (1-3, 4 or 5) for each repetition rate separately. Subsequently, for each rate an ANOVA was performed with repetition number as independent variable and HFB response amplitude as dependent variable, to see if there was a significant difference in HFB-amplitude between repetition numbers. The result of this step was used as an indication of whether there was an influence of previous productions of the same vowel on subsequent productions. Since the slowest production rate only contained three repetitions, the ANOVA was performed on the first two and the last repetition only, for all production rates. Hermes and colleagues (2012b) suggested that a non-linear function in the form of $a*(1/x)+bx+c$ best fitted their results with respect to the shape of the HFB response during finger movements. For visualization of the response profile, we fitted this function with the current data. Furthermore, since those authors found that for finger movements the HFB profile was dependent on movement rate, we compared the HFB response profiles of the three different repetition rates using an ANOVA. The repetition rate was used as independent variable and the HFB response peak amplitude was used as dependent variable (each repetition rate group consisted of the amplitudes

of the first, second and last repetitions combined). Note that also in this step only the first, second and last repetitions were used to allow for comparison across rates.

HFB Response Profiles

Based on previous research (Hermes et al. 2012b; Salari et al. 2018) and on inspection of the data, five models were defined to describe the HFB response profiles of the included electrodes. Electrodes could show (1) high activity for the first vowel production followed by a ‘non-linear decrease’ (NLD) of activity for the remaining productions, (2) high activity for the ‘first production (FP) but none or very little activity for the remaining productions, (3) high activity for the first and last production with a lower response for the productions in between, in the form of a ‘u-shape’ (US), (4) linearly decreasing (LD) activity over productions, or (5) activity could be equally responsive (ER) to all productions. Each electrode was classified as one of these response profiles for each repetition rate separately, by regressing three predictors to the HFB amplitude data of each electrode. Only three predictors were necessary to describe these five profiles as will be explained below. The first predictor models a NLD and a FP profile in a simplified way, with the first peak higher than the other peaks, and the other peaks being more or less equally high. For both the NLD and the FP profile the predictor was [1 0 0 0 0], [1 0 0 0] or [1 0 0] for a five, four and three repetitions trial respectively. If the intercept of the regression was significantly above zero ($\alpha = 0.05$), the whole predictor would be moved up. In that situation, there would be a response present for all repetitions, which differentiates the NLD from the FP profile. The second predictor characterized the US model, (i.e., [1 0 0 0 1], [1 0 0 1] or [1 0 1]). The third predictor represented the LD model, (i.e., [1 .75 .5 .25 0], [1 0.67 0.33 0] or [1 0.5 0]). Note that the slope of this linear predictor was not fixed as the beta and intercept value of the regression determined the slope. The predictor with the highest correlation to the data was chosen as the best fit. Subsequently, we tested if this predictor could significantly explain the amplitude response, based on the beta value from the regression analysis ($\alpha = 0.05$). If an electrode was significant for the best fitting profile (i.e. NDL, FP, US or LD) it was classified as such. If none of the models were significant (and the electrode therefore did not show any difference between the response amplitudes of the repetitions), an electrode was assigned to the ER profile.

We determined, per repetition rate, the percentage of electrodes that belonged to each profile, and evaluated effects of production rate on the number of electrodes per profile. To investigate the presence of an anatomical organization of particular response profiles within the sensorimotor cortex (i.e., whether some profiles are more prominent in specific sensorimotor regions than others), we determined for each electrode if it was classified as the same profile more than once (out of three repetition rates). If so, this profile was considered the most prominent profile for that electrode. We visualized the distribution of these most prominent response profiles on a 3D rendering of the subject’s brain as described in (Hermes et al. 2010; Branco et al. 2018b).

Results

Task Performance and Behavioral/Acoustic Measures

The task was performed well by all subjects, although subject C showed some difficulties during the first run. For subjects A & B, 7.7% (6/78) of the trials were disregarded due to an incorrect response and for subject C this was 35.9% (28/78). For the trials performed accurately (i.e. with the correct number of repetitions), the intended and performed repetition rates did not differ much (see Table 1). Subject A produced the vowels significantly slower than instructed for the fastest production rate, $t(83)=-7.54$, $p<0.001$ and subject C produced them faster for the two fastest repetition rates, $t(43)=2.63$, $p=0.01$ and $t(50)=4.21$, $p<0.001$ respectively.

For subjects A & C there was a significant difference between vowel production durations for the three different repetition rates after Bonferroni correction ($\alpha = 0.05$), $F(2,281)= 5.15$, $p=0.006$ and $F(2,186)= 14.13$, $p<0.001$ respectively, see Table 1. For subject B, the vowel production durations did not differ significantly. Since the difference for subject A is relatively small (only 0.01 second), and there is no significant difference for subject B, these results suggest that there was not a strong overall difference between vowel production duration for the three repetition rates for these two subjects.

The derived behavioral performance measures (sound intensity, lip aperture, lip movement and lip velocity) did not correlate with the brain signal peak amplitudes for most electrodes (see Table 2 for the mean correlation over electrodes) in subjects A & B. In fact, for subject A, none of the electrodes showed a significant correlation to any of the measures. For subject B, only 17.86% (5/28) of the included electrodes showed a significant ($\alpha=0.05$, FDR corrected) correlation of HFB signal amplitude with sound intensity (mean $r=0.36$, $SD=0.09$). For subject C, many electrodes did show a significant correlation with the lip measures; 11.86% (7/59, mean $r=0.26$, $SD=0.05$) with lip position, 74.58% (44/59, mean $r=0.36$, $SD=0.09$) with lip movement, and 47.46% (28/59, mean $r=0.30$, $SD=0.06$) with lip velocity. Note that we correct for these effects in our ECoG analyses, see supplementary Figure S1. This figure shows that most of the signal variability due to for instance sound intensity (see subject B) or lip movement (see subject C) is reduced by the correction we applied and will not have contributed to the results presented in the paper.

Electrode Selection and Peak Timing Models

The models used to select significant electrodes and to determine the peak timings showed an accurate correspondence to the HFB response signals (see Figure 2), with an average of 62% ($SD=12$), 67% ($SD=11$) and 73% ($SD=12$) variance explained for the included electrodes, of subjects A-C respectively.

Average HFB Peak Profile During Vowel Repetitions

In general, the first vowel production of a trial was associated with a larger HFB peak amplitude than subsequent pronunciations (Figure 3). For all subjects, the mean HFB response peak

amplitudes over all significant sensorimotor electrodes differed significantly between repetitions for almost all repetition rates (Table 3, note that we used the first two and the last repetition for all rates). For subject C, there was no significant difference during the 3 repetitions condition.

We investigated whether certain repetition rates were associated with a stronger average decrease in amplitude than other repetition rates. Only subject A showed a significant difference between peak amplitudes over production rates (Figure 4), $F(2,209)=3.27$, $p=0.04$.

Electrode HFB peak profiles

Each included electrode was classified as belonging to one of five response profiles (Figure 5) based on the development of the peak amplitude over repetitions, for each repetition rate separately. In general, the NLD was the most frequent response profile for subjects A & B. For subject C, the US and ER responses were most frequent. None of the response profiles showed a clear anatomical clustering (Figure 6).

We investigated whether there was a general change in the number of electrodes per profile depending on repetition rate (Figure 7). Although no statistical conclusions can be drawn from the results with the current number of subjects, there was an overall trend for an increasing number of ER electrodes with decreasing frequency rate. For subjects A, B and C, respectively, 15.38% (2/13), 20.83% (5/24) and 52.63% (20/38) of all electrodes that showed a repetition effect in the 5 repetitions condition convert to ER in the 3 repetitions condition.

Table 1. Behavioral Performance. *In the top section, the mean and standard deviation of production rate deviations (instructed minus performed rate) over pronunciations are shown in seconds. The asterisk indicates a significant difference between intended and performed production rate. A negative value indicates that vowels were repeated slower than intended. In the lower section, the mean and standard deviation of speech duration is shown per production rate, for all subjects. Discarded trials were not included here.*

Subject	A	B	C
Production rate deviation (s)			
5 reps	-0.06 (0.07)*	0.01 (0.14)	0.04 (0.11)*
4 reps	-0.01 (0.07)	0.03 (0.14)	0.09 (0.16)*
3 reps	0.01 (0.12)	0.04 (0.20)	0.06 (0.18)
Speech durations (s)			
5 reps	0.19 (0.03)	0.27 (0.05)	0.23 (0.07)
4 reps	0.20 (0.03)	0.27 (0.04)	0.21 (0.06)
3 reps	0.20 (0.03)	0.27 (0.04)	0.27 (0.07)
ANOVA	$F(2,281)=5.15$, $p=0.006$	$F(2,281)=1.17$, $p=0.31$	$F(2,186)=14.13$, $p<0.001$

Tongue Movements in Healthy Volunteers

The results from the tongue movement measures indicated that 4 subjects did not show much difference in tongue position over the different repetitions and one subject showed a slightly higher tongue position for the first vowel production compared to subsequent repetitions for the two fastest repetition rates (Figure 8). Whether or not people returned their tongue to the rest position in-between repetitions was quite different over subjects.

Table 2. Correlation between the performance measures and HFB peak amplitudes. *Values are the mean and standard deviation over included electrodes (therefore no p-values are shown).*

Subject	A	B	C
Correlation			
Sound intensity	0.04 (0.08)	0.15 (0.14)	0.02 (0.08)
Lip aperture	0.01 (0.07)	-0.09 (0.08)	0.13 (0.08)
Lip movement	0.05 (0.07)	0.02 (0.07)	0.31 (0.12)
Lip velocity	0.06 (0.06)	0.01 (0.05)	0.21 (0.10)

Discussion

The effect of movement repetition on the sensorimotor HFB response during vowel production was investigated using a simplified speech task with controlled speed of repetition. The HFB signal from high-density electrode grids was evaluated in three epilepsy patients undergoing a surgical procedure for epilepsy diagnosis.

We show that sensorimotor activity related to discrete speech movements is influenced by previous speech movements when spaced a second (or less) apart. Averaged across electrodes, the HFB-response of sensorimotor cortex had a similar amplitude between different production rates but did not show equal amplitudes over the course of repetitions (Figure 3). This was seen for all subjects and all tested production rates (1-1.66 Hz; except for one instance in subject C where the effect was near-significant). The data suggest that the HFB-amplitude is not linearly related to motor output since amplitudes mainly decline non-linearly for repeated vowel productions. The analysis included a correction for the small variations in sound intensity, lip aperture, lip movement and lip velocity, making it unlikely that this finding can be explained by differences in performance over repetitions. However, movements of the tongue could not be measured (discussed below).

The results for speech movements are in agreement with earlier electrophysiological and fMRI data that report a repetition effect for finger movements (Hermes et al. 2012b; Siero et al. 2013). We extend these by showing that complicated movements such as those involved in speech show a decline in HFB response when repeated at a frequency of 1 Hz or higher. Furthermore, our data suggest a tipping-point between 1 and 2 seconds (the production rate at which the HFB amplitude decline disappears) since the repetition effect was still visible for repetitions 1 second apart but was no longer visible after 2 seconds, the time approximately between the last pronunciation of a trial and the first pronunciation of the following trial, as indicated by the recovery

of a high amplitude for each first pronunciation of a trial.

Across electrodes, the non-linear decrease (NLD) profile was dominant for subject A & B. Other response profiles were observed, but less frequently in these subjects. For subject C, the US and ER responses were most frequent. For all subjects, the number of electrodes with an equally responsive (ER) profile tended to increase with decreasing vowel production rate. Considering the earlier discussion on the tipping-point, it could be speculated that different cortical patches in sensorimotor cortex exhibit different tipping points, which would cause more electrodes to display the repetition effect as vowel production rate increases. Note however, that the total number of pronunciations and therewith the number of data points, is different for each repetition rate, which could make the statistical chance to find a particular response profile unequal between repetition rates. Therefore, we cannot fully exclude the possibility that the tipping point effect may be caused partly by an unequal number of data points between repetition rates.

The current results did not indicate a clear anatomical organization with respect to the different response profiles although some clustering seemed to be present (most clearly visible in subject B). Please note that, even though we used HD electrode grids, individual electrodes are still 3 or 4 millimeters apart. Therefore, the spatial sampling is somewhat sparse compared to for instance high field fMRI recordings. Possibly, repeating the experiment with even higher spatial sampling may reveal spatial organization with respect to the repetition effect.

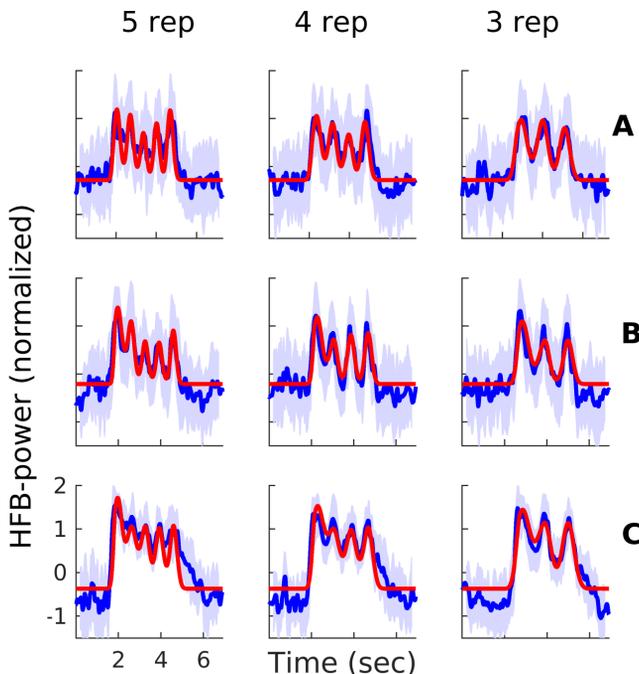


Figure 2. Data and model examples. Each row shows for one subject (A-C) an example of the response of one electrode in blue (average over trials, with shading indicating the standard deviation), with the three different production rates separated in columns (fastest on the left and slowest on the right). In red, the model that best fitted that specific electrode's response is shown. On the x-axis, time is indicated in seconds after the first cue. On the y-axis, the normalized HFB-power is indicated.

Table 3. HFB peak amplitudes over repetitions. *The ANOVA results per production rate and per subject, testing for differences between HFB peak amplitudes over repetitions. The dependent variable was the mean (over included electrodes) HFB peak amplitudes and the independent variable was the repetition number. Significance indicates that the HFB peak amplitudes are significantly different between repetitions.*

Subject	A	B	C
5 reps	$F(2,59)=20.61, p<0.001$	$F(2,62)=17.28, p<0.001$	$F(2,30)=4.69, p=0.02$
4 reps	$F(2,73)=25.63, p<0.001$	$F(2,68)=24.73, p<0.001$	$F(2,45)=6.26, p<0.001$
3 reps	$F(2,71)=19.49, p<0.001$	$F(2,74)=28.43, p<0.001$	$F(2,60)=3.03, p=0.06$

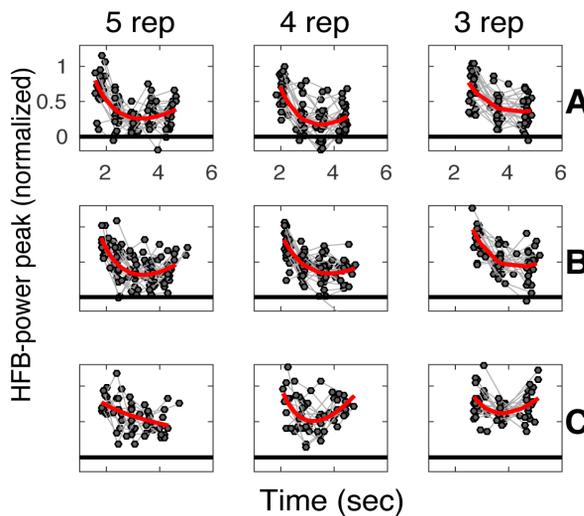


Figure 3. The average HFB profile for different vowel repetition rates. *Each panel shows for one subject (A-C) the average profile of the HFB peak response for each vowel production averaged over included electrodes (marked by gray circles, pronunciations of the same trial are connected by gray lines). The three different production rates are separated in columns (fastest on the left and slowest on the right). A non-linear line, shown in red, was fitted to the data for visualization and for comparison with previous studies of hand movements (Hermes et al. 2012b). On the x-axis, time is indicated in seconds after the first cue and on the y-axis the normalized HFB-power is indicated.*

Neural Underpinnings of the Repetition Effect

There are several phenomena that may account for the decrease in HFB-power observed with repeated speech movements. It may be speculated that some articulators which we did not correct for, moved more for the first pronunciation than for subsequent pronunciations. Analysis of tongue movements during the same task in healthy volunteers revealed quite constant tongue positions

over repetitions within subjects but also revealed variations in tongue movements between subjects, ranging from full contraction and relaxation for each repetition to a fixed tongue position throughout repetitions (Figure 8). Previous research has suggested that not only articulator position but also features related to articulator movements (such as velocity) are represented in the sensorimotor cortex (Conant et al. 2018). In case the tongue may not return to its rest position between repetitions, the first utterance may be associated with more activation than the subsequent productions as the articulator movement is then largest for the first pronunciation and smaller for the subsequent ones. This may also explain why the last repetition sometimes showed an increase in activity compared to its predecessor(s) as the tongue has to return to its rest position. Furthermore, it can be speculated, that between phonemes the musculature used for the production is not fully at rest (in anticipation of the next production), even if the articulator position between phonemes is close to the rest position. In this case, one could see the full sequence of repetitions as one, albeit complex, movement with an onset and an offset. Since various reports have shown a neural response at movement offset (Ball et al. 2008; Hermes et al. 2012b; Salari et al. 2018), neural activity at the end of a sequence may also be attributed to the movement towards full rest. However, this cannot explain all the found response profiles. Hence, our findings cannot be fully explained by differences in tongue movements between repetitions and are therefore also in line with the existence of a non-linearity between motor output and neural activity during repeated speech-movements that are spaced closely apart.

Since HFB power is thought to be associated with neural firing (Manning et al. 2009; Miller et al. 2009; Ray and Maunsell 2011), a decrease in HFB power as observed here may suggest that fewer neurons are involved in subsequent motor acts, (see Hermes et al. 2012b for a similar interpretation), or that the same neurons fire less frequently. Indeed, repetition suppression effects

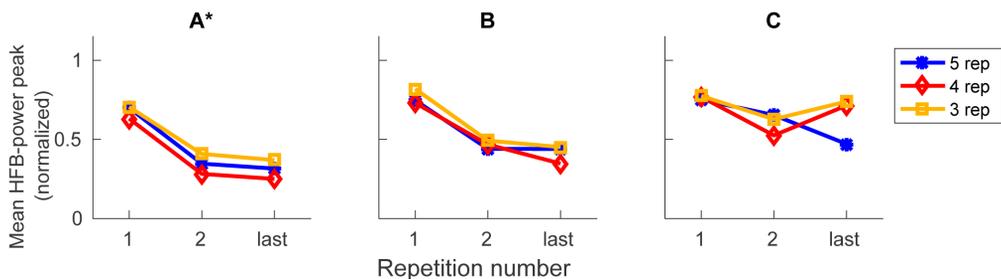


Figure 4. Comparison of HFB profiles between different production rates. Each panel shows for one subject (A-C), the HFB amplitude per repetition number, averaged over included electrodes and over trials. The three different production rates are indicated in blue, red and yellow (for the first two and the last repetition). On the x-axis, the repetition number is indicated and on the y-axis the normalized HFB-power is indicated. A significant difference between conditions is indicated by an asterisk.

have been found for other modalities than motor execution. Suppression during repeated visual stimulation has for example been attributed to a reduction of neural excitability for repeated stimuli (see Grill-Spector et al. 2006 for an interesting discussion on the possible mechanisms behind a reduction of neural activity for repeated stimuli and the possible function this may have).

Furthermore, repetition suppression effects for repeated speech have also been found in other areas than the sensorimotor cortex and may be involved in motor planning. Previous fMRI research in the left posterior inferior frontal gyrus, has shown for instance, a repetition suppression effect which is related to the degree of shared phonological features (such as voicing or manner of articulation) over the course of repeated words (Okada et al. 2018). This suggests that similar phonological features during speech reduce activity in motor planning areas. It would be interesting to see if such motor planning effects of similar phonological features is related to the repetition effect we observed in the sensorimotor cortex as it has been suggested that repetition effects in some areas may affect the activity in other areas (Grill-Spector et al. 2006).

Furthermore, even though in the current study we focused on sensorimotor cortex activity, other areas such as the supplementary motor area (SMA), cerebellum, basal ganglia and premotor cortex, which are connected to the motor cortex, have been suggested to play an important role in the timing of speech production (Kotz and Schwartz 2010). It would therefore be interesting to investigate the role of those areas during repeated speech production and to see if they have an influence on the repetition effect.

Neural Underpinnings of Different Response Types

Although further research is needed to better understand the different response types we found, we may speculate about the possible underlying mechanisms. There are multiple theories on the mechanisms behind repetition suppression that may explain the current results. For instance, increased influx of potassium ions over the course of repetitions may lead to hyperpolarization of the cell membrane, causing a reduction in neural firing. If this effect is asymptotic, this may lead to a non-linear decrease. Another theory suggests that only neurons that are most specific to the task continue firing over repetitions. It could be that in some areas the number of task-specific neurons is higher than in others, which may lead to the different response types. If most of the neurons are task-specific, it would be likely that the responses are equal over repetitions (ER). If the ratio between task-specific neurons and task-unspecific neurons is high, the number of firing neurons may decrease initially and stabilize at some point, leading to a non-linear decrease (NLD or FP). With a lower ratio, the decrease may be linear, as the number of neurons that can stop firing is larger. As discussed earlier, some areas may be related to movements in two directions (Fetz et al. 1980; Soso and Fetz 1980), which may explain the u-shape response type (US), as at the end of the trial the articulators are likely to return to rest position, see for instance the tongue position data in Figure 8. This may result in increased neural firing at the end of the trial. Besides these theories on the neural underpinnings of repetition suppression, there may be an alternative explanation for a higher response for the first vowel production. Some parts of the sensorimotor cortex may be involved in the planning of a motor sequence (Tanji and Evarts 1976) and it may be speculated that this could lead to more neural activity for the first production (i.e. beginning of a rhythmic sequence) or to only neural activity at the beginning of the sequence.

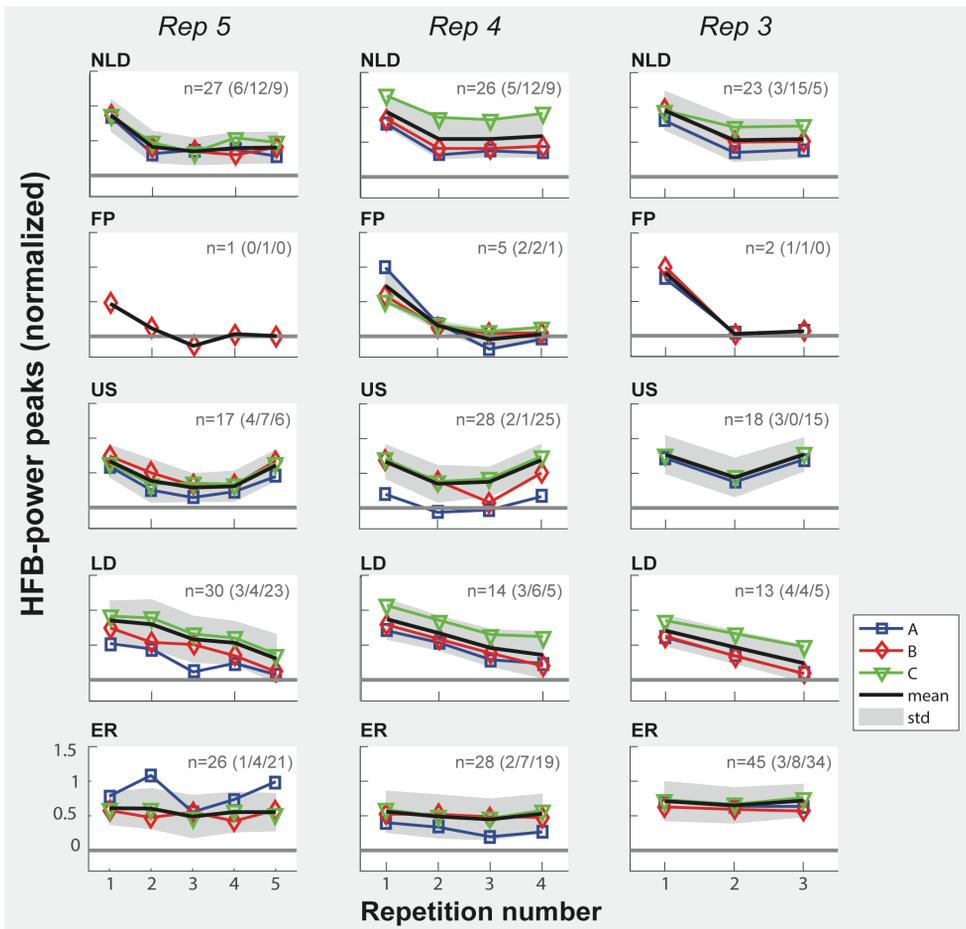


Figure 5. Representation of all response profiles. Responses were ‘non-linear decreasing’ (NLD) ‘first production responsive only’ (FP), ‘u-shaped’ (US), ‘linearly decreasing’ (LD) or ‘equally responsive’ (ER). Each row indicates the average response (per subject in color and for all subjects in black) of all electrodes that belonged to one profile. The number of electrodes on which the mean response was based is indicated by ‘n’ and the number of electrodes each subject contributed is indicated by the numbers in brackets for subject A-C respectively. Columns separate the different production rates. The standard deviation of the response over subjects is shown by shading. On the x-axis, the repetition number is indicated and on the y-axis the normalized HFB-power is indicated.

Implications for Neural Based Speech Decoding

Our results are highly relevant for the development of sensorimotor-speech BCIs: systems that aim to decode (attempted) speech from sensorimotor brain signals. Classification of speech sounds based on sensorimotor activity has been shown before (Kellis et al. 2010; Brumberg et al. 2011;

Mugler et al. 2014; Herff et al. 2015; Ramsey et al. 2018), but accuracy levels and degrees of freedom do not meet the standards for home-use by patients. It has been suggested however, that these systems are likely to benefit from the use of high-density electrode grids (Kellis et al. 2016; Ramsey et al. 2018). Classification of sensorimotor signals may also benefit from taking linguistic structures, such as syntax or likely word combinations, into consideration by incorporating a language corpus to the predictions (Herff et al. 2015, 2017).

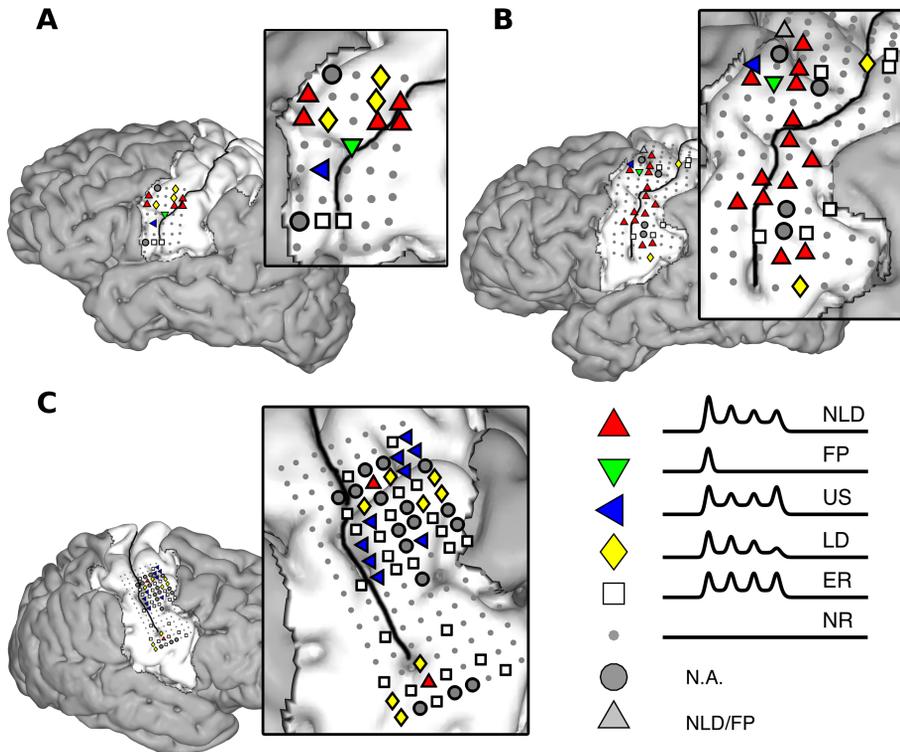


Figure 6. Electrode localization. For each subject, the locations of all analyzed electrodes are indicated. Small gray dots indicate non-responsive (NR) electrodes. The other symbols indicate, per electrode, the most prominent response profile, which was determined by classifying the HFB responses for each production rate into one of five response profiles (for abbreviation clarification of the response profiles see Figure 5). If an electrode was classified as the same response profile for at least two out of the three production rates, this was considered the most prominent response profile. Gray circles (see N.A. in legend) indicate electrodes that did not have a prominent response (i.e. where the responses were different for each production rate). The difference between the NLD and FP response predictor was an intercept difference and therefore light gray triangles (see NLD/FP in legend) indicate electrodes which had different responses for each repetition rate but which were classified once as NLD and once as FP.

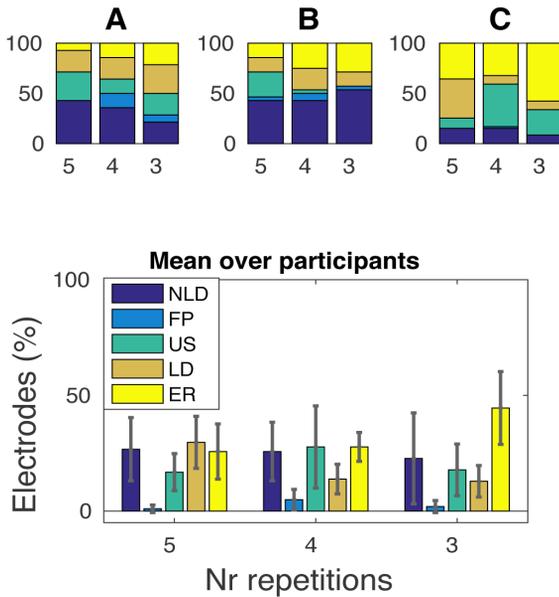


Figure 7. Number of electrodes for each profile. In the upper panels, the percentage of included electrodes belonging to a profile (indicated by color) is shown for all subjects and production rates (indicated on the x-axis). On the y-axis, the percentage of electrodes is indicated. For subject A-C, a total number of 14, 28 and 59 electrodes were significantly active, respectively, which corresponds to the 100% value for each subject. The lower panel shows the weighted mean and standard deviation values over subjects.

We postulate that a third factor needs to be taken into account for optimal decoding accuracy: previous speech movements. Since sensorimotor-speech BCIs try to find specific patterns in the brain signals that can consistently be linked to a specific sound, and use this information to determine which sound the user made (or tried to make), any effect of previous (attempted) utterances on the brain signal of a current utterance is important information. Sensorimotor-speech BCIs may therefore be improved when information about previously spoken sounds is incorporated in the decoding pipeline. The current study provides a method for creating models of HFB profiles related to vowel repetitions. Models like these may be used for creating a library of models related to variations in brain activity patterns associated with speech production and may potentially improve speech classification. It will be crucial to extend the current findings to more real-life application scenarios of sensorimotor-speech BCIs, and to investigate whether the results can be generalized to natural speech circumstances such as repeating the same phoneme within a word or over the course of words. Furthermore, since these BCI systems are intended for paralyzed subjects, it is essential to investigate if also repeated covert/attempted speech is associated with similar phenomena.

Limitations & Future Work

One of the limitations of the current study is the small number of subjects. Yet all subjects show similar results (decreased activity for repeated vowel production) across all investigated production rates (except for one instance in subject C where the effect was near-significant) and our findings do correspond to that of previous studies for finger movements. Also, a larger range of production rates could have been more informative, notably to determine the tipping point for HFB response

recovery. Third, we did not record articulator positions directly (except for the lips) and could therefore not correct for all variations in motor output. Fourth, in the current study we did not control for a possible effect of auditory stimulation on the cortical responses (by the subjects hearing their own voice during the task). However, since previous studies of other repeated movements (not involving speech or auditory stimulation) have shown similar results as the current study (e.g. Hermes et al. 2012b), we argue that it is likely that the repetition effect is more a sensorimotor cortex effect related to movements than to auditory stimulation. Finally, from our study it is not possible to determine whether the repetition effect is specific for the same phoneme, or could generalize to different phonemes following one another. This issue clearly warrants further investigation, as it is relevant for decoding speech where different phonemes are produced in sequence.

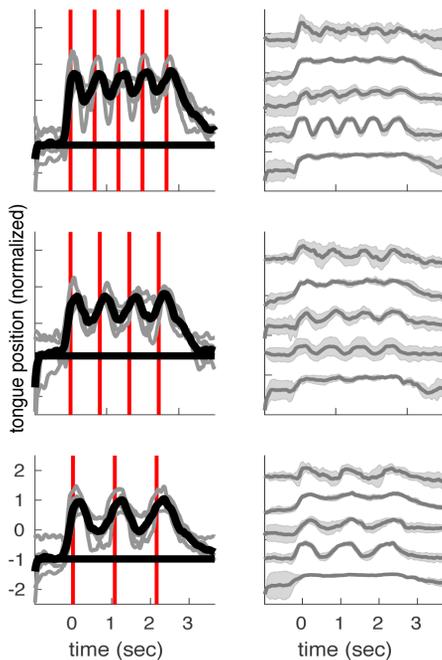


Figure 8. Tongue positions. *The average tongue position height (normalized) of five healthy volunteers is shown on the y-axis. The part of the tongue that moved the most with the task is shown. On the x-axis, time is indicated in seconds. Red vertical lines indicate voice onsets. The three different repetition rates are shown from top to bottom. The data was corrected for differences in pronunciation timing and was subsequently averaged per subject. Gray lines indicate the average tongue position of the individual volunteers (shading indicates the standard deviation) and the black line in the left panel represents the average thereof.*

Conclusion

We show that neural activity related to discrete repeated speech movements is influenced by previous speech movements spaced a second or less apart. The most prominent response profile for repeated speech movements is a non-linear decrease of neural activity over repetitions. These findings are of importance for the development of communication-brain-computer interfaces that use decoding of (overt or covert) speech.

Acknowledgements

The authors would like to thank the participants, the staff of the clinical neurophysiology department and the neurosurgeons for their contribution and Etske Ooijevaar for her help with the ultrasound measurements.

References

- American Congress of Rehabilitation Medicine (1995) Recommendations for use of uniform nomenclature pertinent to patients with severe alterations in consciousness. *Arch Phys Med Rehabil* 76:205–209. doi: 10.1016/S0003-9993(95)80031-X
- Articulate Instruments Ltd., (2008) Ultrasound stabilisation headset users manual: revision 1.4. Articulate Instruments Ltd, Edinburgh
- Articulate Instruments Ltd., (2012) Articulate assistant advanced user guide: version 2.14. Articulate Instruments Ltd, Edinburgh
- Ball T, Demandt E, Mutschler I, et al (2008) Movement related activity in the high gamma range of the human EEG. *NeuroImage* 41:302–310. doi: 10.1016/j.neuroimage.2008.02.032
- Bleichner MG, Jansma JM, Salari E, et al (2015) Classification of mouth movements using 7 T fMRI. *J Neural Eng* 12:066026. doi: 10.1088/1741-2560/12/6/066026
- Boersma P (2002) Praat, a system for doing phonetics by computer. *Glott Int* 5.
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–332. doi: 10.1038/nature11911
- Branco MP, Freudenburg ZV, Aarnoutse EJ, et al (2018a) Optimization of sampling rate and smoothing improves classification of high frequency power in electrocorticographic brain signals. *Biomed Phys Eng Express* 4:045012. doi: 10.1088/2057-1976/aac3ac
- Branco MP, Gaglianesi A, Glen DR, et al (2018b) ALICE: A tool for automatic localization of intra-cranial electrodes for clinical and high-density grids. *J Neurosci Methods* 301:43–51. doi: 10.1016/j.jneumeth.2017.10.022
- Brumberg JS, Wright EJ, Andreasen DS, et al (2011) Classification of Intended Phoneme Production from Chronic Intracortical Microelectrode Recordings in Speech-Motor Cortex. *Front Neurosci* 5. doi: 10.3389/fnins.2011.00065
- Bruns A (2004) Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? *J Neurosci Methods* 137:321–332. doi: 10.1016/j.jneumeth.2004.03.002
- Conant DF, Bouchard KE, Leonard MK, Chang EF (2018) Human sensorimotor cortex control of directly-measured vocal tract movements during vowel production. *J Neurosci* 2382–17. doi: 10.1523/JNEUROSCI.2382-17.2018
- Coudé G, Ferrari PF, Rodà F, et al (2011) Neurons Controlling Voluntary Vocalization in the Macaque Ventral Premotor Cortex. *PLOS ONE* 6:e26822. doi: 10.1371/journal.pone.0026822
- Crone NE, Hao L, Hart J, et al (2001) Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology* 57:2045–2053. doi: 10.1212/WNL.57.11.2045
- Crone NE, Miglioretti DL, Gordon B, Lesser RP (1998) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* 121:2301–2315. doi: 10.1093/brain/121.12.2301
- Donoghue JP, Sanes JN, Hatsopoulos NG, Gaál G (1998) Neural Discharge and Local Field Potential Oscillations in Primate Motor Cortex During Voluntary Movements. *J Neurophysiol* 79:159–173
- Fetz EE, Finocchio DV, Baker MA, Soso MJ (1980) Sensory and motor responses of precentral cortex cells during comparable passive and active joint movements. *J Neurophysiol* 43:1070–1089
- Georgopoulos, Ashe J, Smyrnis N, Taira M (1992) The Motor Cortex and the Coding of Force. *Science* 256:1692–1695
- Georgopoulos, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J Neurosci* 2:1527–1537
- Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14–23. doi: 10.1016/j.tics.2005.11.006
- Herff C, Heger D, de Pesters A, et al (2015) Brain-to-text: decoding spoken phrases from phone representations in the brain. *Neural Technol* 9:217. doi: 10.3389/fnins.2015.00217
- Herff C, Pesters A de, Heger D, et al (2017) Towards Continuous Speech Recognition for BCI. *SpringerLink* 21–29. doi: 10.1007/978-3-319-57132-4_3
- Hermes D, Miller KJ, Noordmans HJ, et al (2010) Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J Neurosci Methods* 185:293–298. doi: 10.1016/j.jneumeth.2009.10.005
- Hermes D, Miller KJ, Vansteensel MJ, et al (2012a) Neurophysiologic correlates of fMRI in human motor cortex. *Hum Brain Mapp* 33:1689–1699. doi: 10.1002/hbm.21314
- Hermes D, Siero JCW, Aarnoutse EJ, et al (2012b) Dissociation between Neuronal Activity in Sensorimotor Cortex and Hand Movement Revealed as a Function of Movement Rate. *J Neurosci* 32:9736–9744. doi:

10.1523/JNEUROSCI.0357-12.2012

- Hermes DJ (1990) Vowel-onset detection. *J Acoust Soc Am* 87:866–873. doi: 10.1121/1.398896
- Kellis S, Miller K, Thomson K, et al (2010) Decoding spoken words using local field potentials recorded from the cortical surface. *J Neural Eng* 7:056007. doi: 10.1088/1741-2560/7/5/056007
- Kellis S, Sorensen L, Darvas F, et al (2016) Multi-scale analysis of neural activity in humans: Implications for micro-scale electrocorticography. *Clin Neurophysiol* 127:591–601. doi: 10.1016/j.clinph.2015.06.002
- Kennerley SW, Diedrichsen J, Hazeltine E, et al (2002) Callosotomy patients exhibit temporal uncoupling during continuous bimanual movements. *Nat Neurosci* 5:376–381. doi: 10.1038/nn822
- Kotz SA, Schwartz M (2010) Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cogn Sci* 14:392–399. doi: 10.1016/j.tics.2010.06.005
- Logothetis NK, Pauls J, Augath M, et al (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157. doi: 10.1038/35084005
- Manning JR, Jacobs J, Fried I, Kahana MJ (2009) Broadband shifts in LFP power spectra are correlated with single-neuron spiking in humans. *J Neurosci Off J Soc Neurosci* 29:13613. doi: 10.1523/JNEUROSCI.2041-09.2009
- Miall RC, Ivry R (2004) Moving to a different beat. *Nat Neurosci* 7:1025–1026. doi: 10.1038/nn1004-1025
- Miezin FM, Maccotta L, Ollinger JM, et al (2000) Characterizing the Hemodynamic Response: Effects of Presentation Rate, Sampling Procedure, and the Possibility of Ordering Brain Activity Based on Relative Timing. *NeuroImage* 11:735–759. doi: 10.1006/nimg.2000.0568
- Miller KJ, Leuthardt EC, Schalk G, et al (2007) Spectral Changes in Cortical Surface Potentials during Motor Movement. *J Neurosci* 27:2424–2432. doi: 10.1523/JNEUROSCI.3886-06.2007
- Miller KJ, Sorensen LB, Ojemann JG, Nijss M den (2009) Power-Law Scaling in the Brain Surface Electric Potential. *PLOS Comput Biol* 5:e1000609. doi: 10.1371/journal.pcbi.1000609
- Moran DW, Schwartz AB (1999) Motor Cortical Representation of Speed and Direction During Reaching. *J Neurophysiol* 82:2676–2692
- Mugler EM, Patton JL, Flint RD, et al (2014) Direct classification of all American English phonemes using signals from functional speech motor cortex. *J Neural Eng* 11:035015. doi: 10.1088/1741-2560/11/3/035015
- Okada K, Matchin W, Hickok G (2018) Phonological Feature Repetition Suppression in the Left Inferior Frontal Gyrus. *J Cogn Neurosci* 1–9. doi: 10.1162/jocn_a_01287
- Pei X, Leuthardt EC, Gaona CM, et al (2011) Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *NeuroImage* 54:2960–2972. doi: 10.1016/j.neuroimage.2010.10.029
- Penfield W, Boldrey E (1937) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain J Neurol* 60:389–443. doi: 10.1093/brain/60.4.389
- Posner JB, Plum F, Saper CB, Schiff N (2007) Plum and Posner's Diagnosis of Stupor and Coma. Oxford University Press, USA
- Ramsey NF, Salari E, Aarnoutse EJ, et al (2018) Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage*. doi: 10.1016/j.neuroimage.2017.10.011
- Ray S, Maunsell JHR (2011) Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLOS Biol* 9:e1000610. doi: 10.1371/journal.pbio.1000610
- Salari E, Freudenburg ZV, Vansteensel MJ, Ramsey NF (2018) Spatial-Temporal Dynamics of the Sensorimotor Cortex: Sustained and Transient Activity. *IEEE Trans Neural Syst Rehabil Eng* 26:1084–1092. doi: 10.1109/TNSRE.2018.2821058
- Schaal S, Sternad D, Osu R, Kawato M (2004) Rhythmic arm movement is not discrete. *Nat Neurosci* 7:1136–1143. doi: 10.1038/nn1322
- Siero JC, Hermes D, Hoogduin H, et al (2013) BOLD Consistently Matches Electrophysiology in Human Sensorimotor Cortex at Increasing Movement Rates: A Combined 7T fMRI and ECoG Study on Neurovascular Coupling. *J Cereb Blood Flow Metab* 33:1448–1456. doi: 10.1038/jcbfm.2013.97
- Siero JC, Hermes D, Hoogduin H, et al (2014) BOLD matches neuronal activity at the mm scale: A combined 7 T fMRI and ECoG study in human sensorimotor cortex. *NeuroImage* 101:177–184. doi: 10.1016/j.neuroimage.2014.07.002
- Smith E, Delargy M (2005) Locked-in syndrome. *BMJ* 330:406–409. doi: 10.1136/bmj.330.7488.406
- Soltysik DA, Peck KK, White KD, et al (2004) Comparison of hemodynamic response nonlinearity across primary cortical areas. *NeuroImage* 22:1117–1127. doi: 10.1016/j.neuroimage.2004.03.024
- Soso MJ, Fetz EE (1980) Responses of identified cells in postcentral cortex of awake monkeys during comparable active and passive joint movements. *J Neurophysiol* 43:1090–1110
- Spencer RMC, Zelaznik HN, Diedrichsen J, Ivry RB (2003) Disrupted Timing of Discontinuous But Not Continuous Movements by Cerebellar Lesions. *Science* 300:1437–1439. doi: 10.1126/science.1083661

- Tanji J, Evarts EV (1976) Anticipatory activity of motor cortex neurons in relation to direction of an intended movement. *J Neurophysiol* 39:1062–1068
- Towle VL, Yoon H-A, Castelle M, et al (2008) ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain* 131:2013–2027. doi: 10.1093/brain/awn147
- Truccolo W, Friehs GM, Donoghue JP, Hochberg LR (2008) Primary Motor Cortex Tuning to Intended Movement Kinematics in Humans with Tetraplegia. *J Neurosci* 28:1163–1178. doi: 10.1523/JNEUROSCI.4415-07.2008
- Wang W, Chan SS, Heldman DA, Moran DW (2007) Motor Cortical Representation of Position and Velocity During Reaching. *J Neurophysiol* 97:4258–4270. doi: 10.1152/jn.01180.2006
- Wolpaw JR, Birbaumer N, McFarland DJ, et al (2002) Brain–computer interfaces for communication and control. *Clin Neurophysiol* 113:767–791. doi: 10.1016/S1388-2457(02)00057-3

Appendix

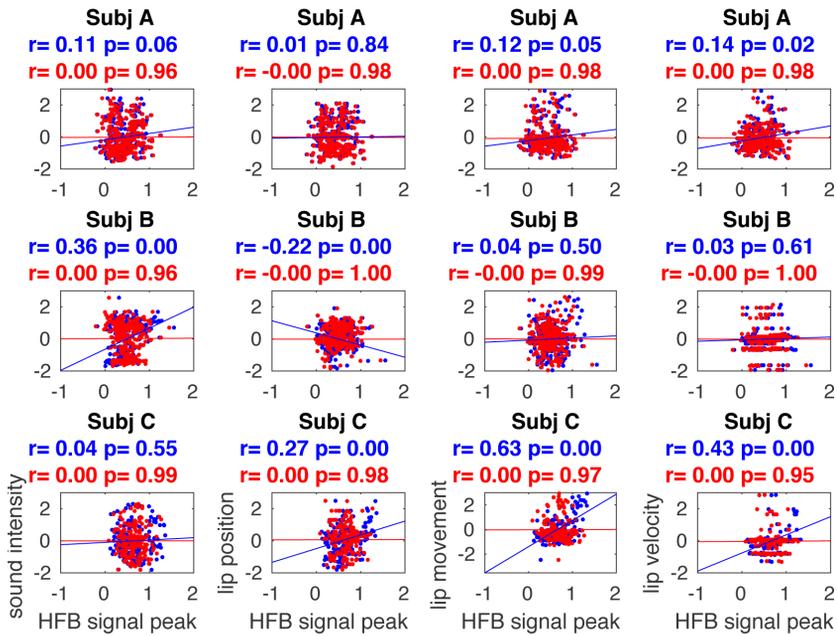


Figure S1. Influence of behavioral measure corrections. Correlation, for subject A-C, between four behavioral measures (sound intensity, lip position, lip movement and lip velocity) with the normalized brain signal peak amplitudes, averaged over electrodes, before (blue) and after correction (red) for behavioral measures. On the x-axis, the HFB signal peak amplitude is indicated and on the y-axis the behavioral measure. The correlation value (r) and significance value (p) are indicated above each plot in the corresponding color.





Chapter 3

*“Of science and the human heart, there is no limit.
There is no failure here sweetheart, just when you quit.”*

Chapter 3

The Influence of Prior Pronunciations on Sensorimotor Cortex Activity Patterns During Vowel Production

E. Salari, Z.V. Freudenburg, M.J. Vansteensel, N.F. Ramsey

Published Article: Journal of Neural Engineering, September 2018

Abstract

Background: In recent years, brain-computer interface (BCI) systems have been investigated for their potential as a communication device to assist people with severe paralysis. Decoding speech sensorimotor cortex activity is a promising avenue for the generation of BCI control signals, but is complicated by variability in neural patterns, leading to suboptimal decoding. We investigated whether neural pattern variability associated with sound pronunciation can be explained by prior pronunciations and determined to what extent prior speech affects BCI decoding accuracy. **Methods:** Neural patterns in speech motor areas were evaluated with electrocorticography in five epilepsy patients, who performed a simple speech task that involved pronunciation of the /i/ sound, preceded by either silence, the /a/ sound or the /u/ sound. **Results:** The neural pattern related to the /i/ sound depends on previous sounds and is therefore associated with multiple distinct sensorimotor patterns, which is likely to reflect differences in the movements towards this sound. We also show that these patterns still contain a commonality that is distinct from the other vowel sounds (/a/ and /u/). Classification accuracies for the decoding of different sounds do increase, however, when the multiple patterns for the /i/ sound are taken into account. Simply including multiple forms of the /i/ vowel in the training set for the creation of a single /i/ model performs as well as training individual models for each /i/ variation. **Discussion:** Our results are of interest for the development of BCIs that aim to decode speech sounds from the sensorimotor cortex, since they argue that a multitude of cortical activity patterns associated with speech movements can be reduced to a basis set of models which reflect meaningful language units (vowels), yet it is important to account for the variety of neural patterns associated with a single sound in the training process.

Keywords: Coarticulation, Sensorimotor Cortex, Speech, Brain-Computer Interface, ECoG

Introduction

People who suffer from severe paralysis can sometimes lose the ability to speak or communicate while retaining adequate levels of cognitive functioning (American Congress of Rehabilitation Medicine 1995; Smith and Delargy 2005; Posner et al. 2007). This condition is called locked-in syndrome (LIS). People with LIS often rely on assistive communication technology to interact with their family and caretakers. An example of such technology is a brain-computer interface (BCI) system (Wolpaw et al. 2002). BCIs record brain signals and translate specific features of the recorded signal into a control signal for a computer or other effector. As such, these systems aim to provide people with paralysis the ability to, for instance, write text on a screen or produce synthesized speech, without the involvement of any muscle activity (see for instance Vansteensel et al. 2016 for an example of a fully implantable BCI system for home use).

Recent years have seen an increasing interest in decoding brain activity related to speech movements for the control of BCIs. Several studies using implanted electrodes have shown, for instance, that sounds (isolated or within words) can be classified from the sensorimotor cortex (accuracy for 4 sounds: $71.9 \pm 8.8\%$, Ramsey et al. 2018; accuracy for a complete set of American English phonemes $20.4 \pm 9.8\%$, Mugler et al. 2014). Also, when speech is not actually executed but *attempted* by someone with paralysis, classification of 38 American English phonemes was possible with an average accuracy of 21% (Brumberg et al. 2011). Also in natural spoken speech, it is possible to reconstruct what people said, based on brain signals alone (word classification up to 75% in a set of 10 words, Herff et al. 2015). However, these systems are far from being a usable application for people from the target population since the accuracy levels and degrees of freedom of speech motor BCIs remain suboptimal.

A possible explanation for the suboptimal performance of current speech motor BCIs is that most decoders do not explicitly take into account that there may be variations in the brain activity patterns related to sound pronunciation and generally base classification of each individual sound on a single activity pattern. However, for speech recognition applications that are based on sound recordings (Lee 1990) or facial muscle movements (Schultz and Wand 2010), it has been shown that taking context (previous and future phonemes) into consideration, substantially improves speech recognition. The reason for this might be that articulator placement for a particular sound depends on the sound(s) pronounced before and/or after the spoken sound, a phenomenon called coarticulation (Rietveld and van Heuven 2001; Davis 2015). Similarly, the movements the articulators make to create a particular sound are likely to depend on prior sounds, since the articulators move from different starting positions. Because articulator movement is largely controlled by the sensorimotor cortex of the brain (Penfield and Boldrey 1937; Crone et al. 1998; Towle et al. 2008; Bouchard et al. 2013), variations in articulator positions for the same sound, or in their movements towards a certain position, are likely to be associated with variations in sensorimotor brain activity patterns. Indeed, there is evidence for a representation of coarticulation in the sensorimotor cortex in that, for single electrodes, the high-frequency band (HFB; >50 Hz) power amplitude varies with different articulator positions for the same sound (Bouchard and Chang 2014; Chartier et al. 2018). Also, we have previously shown that repeatedly pronouncing the same sound is not associated with equal neural activity in the sensorimotor cortex if the sounds

are spaced a second or less apart, suggesting that previous sounds influence sensorimotor cortex activity (Salari et al. 2018b). It is currently unclear, however, how coarticulation and the movement from one sound to another, affect our ability to distinguish a certain sound from other sounds, using neural signals. Also, it remains to be determined if the incorporation of multiple brain patterns for a single sound in a decoder can improve classification accuracies. Clearly, answers to these questions are of great interest for the development of motor speech BCIs since they may improve the classification performance and thereby daily life usability of such systems.

In the current study, we confirm that sensorimotor cortex neural activity patterns that are related to the pronunciation of one specific sound, and likely the movements towards that sound (the /i/ vowel in this study), depends on the preceding sound. In addition, we investigate the consequences of these different /i/-related activity patterns for the decoding of different sounds based on sensorimotor cortex neural signals. We use electrocorticography (ECoG) to record neural signals since this technique encompasses a combination of high temporal resolution and high spatial specificity (Siero et al. 2014), both particularly relevant for speech related research (Kellis et al. 2016; Ramsey et al. 2018). We focus on the 75-135 Hz power changes in the sensorimotor cortex, which are known to be closely related to speech and articulator movements, especially in the ventral areas of the sensorimotor cortex (Crone et al. 1998; Miller et al. 2007; Bouchard et al. 2013), and which have been associated with underlying neural firing (Miller et al. 2007; Manning et al. 2009; Ray and Maunsell 2011).

Methods

Participants

Five patients (age 14-41; 3 females) participated in this study, while they were treated for epilepsy in the University Medical Center Utrecht. Subdural ECoG electrodes were placed for clinical reasons. In two participants, the coverage of the clinical grids included the area of interest, the ventral sensorimotor cortex (vSMC; left hemisphere in one participant, right in the other participant). The other three participants (all adults) gave informed consent for the placement of an extra high-density (HD) electrode grid over this area for research purposes only (left hemisphere in 2 participants, right in 1 participant). Electrodes of the clinical grids were 2.3 mm in diameter and were spaced 10 mm apart. HD-grid electrodes were 1 or 1.17 mm in diameter and spaced 3 or 4 mm apart. For the participants with HD-grid coverage we only used the HD electrodes for the current analysis.

This study is in accordance with the Declaration of Helsinki (2013) and was approved by the ethics committee of the University Medical Center Utrecht. All participants gave written informed consent before participation.

Task

Participants were visually cued to produce one of two vowels (the /u/ or /a/), or to remain silent, for 1000 ms, and then make a smooth transition from the current sound (or silence) into a new sound (the /i/ vowel), without a break between the two sounds, and hold the /i/ for 1000 ms. We chose these vowels since they are phonetically well described and are known to involve different

articulator placements (Booij 1999). Visual cues were presented on a computer screen, which was placed approximately 1 m from the subject. A trial started with the visual presentation of either the letters ‘oe’ or ‘aa’, which, in Dutch, correspond to the vowels /u/ and /a/, respectively, or with a ‘-’ sign, which indicated to remain silent. This was followed by the presentation of the letters ‘ie’, corresponding to the /i/ sound. Subjects were instructed to produce the sound that was presented on the screen for as long as it was visible. Trials were interleaved with inter-trial intervals of 3000 ms, during which a fixation cross was presented on the screen. Conditions (/u/ followed by /i/, /a/ followed by /i/, and silence followed by /i/) were randomly repeated 30 times, divided over two sessions (except for subject B, who completed only 1 session and therefore performed 15 trials per condition). We also included 20 rest trials of 2000 ms (10 for subject B), interleaved with the other trials, in which the subjects remained silent.

Data Acquisition & Preprocessing

The recorded sampling frequency of the ECoG data was either 512 Hz, 2048 Hz (both Micromed, Treviso, Italy), or 2000Hz (Blackrock Microsystems LLC, Salt Lake City, USA), similar as in Salari et al. (2018a). For the subjects who performed two runs, we concatenated the data of the two runs. Audio was recorded by microphones in the patient’s room and synchronized with the brain signals. Unless mentioned otherwise, data analysis was performed using Matlab software (The Mathworks, Inc., Natick, MA, USA).

Electrodes over the sensorimotor cortex were identified by visual inspection of a 3D rendering of the subjects preoperative MRI scan that was coregistered with a post-operative CT scan (Hermes et al. 2010; Branco et al. 2018b). Electrodes with a flat or noisy signal, or with signal that was considered epileptic by a neurologist, were removed from further analysis (see Salari et al. 2018a for details). For the remaining electrodes (n=25, 18, 36, 114 and 122 for subjects A-E, respectively), line noise removal and common-average re-referencing was applied. All mention of electrodes below pertain to only the remaining electrodes. Subsequently, for each of the electrodes the HFB (75-135 Hz) power trace was calculated with a Gabor wavelet (Bruns 2004), similar as in Salari et al. (2018a), and this signal was subsequently smoothed for noise reduction with a moving average window of 0.1 seconds, centered around each sample. We choose this window since it is within an optimal range for phoneme classification (Branco et al. 2018a), while at the same time preserving neural dynamics (Salari et al. 2018b). We normalized the signals over runs by taking, per run, the average HFB amplitude (over all electrode time series), subtracting this from all electrode time series and dividing the result by the standard deviation. The voice on- and offsets were first automatically determined from the acoustic signal using an automated vowel detection algorithm (Hermes 1990) and, where necessary they were corrected by auditory and visual inspection using PRAAT software (Boersma 2002). In addition, we determined the voice transition from one sound to another by visual and auditory inspection of the acoustic signal.

Electrode Selection

For every sensorimotor electrode, we determined whether it was significantly related to the pronunciation of one or more of the sounds, by computing per sound production (the three different

/i/ productions and the /u/ and /a/ production) the mean HFB response amplitude in a window of 0.3 seconds around sound onset. The HFB response amplitude during sound production was contrasted with the mean HFB amplitude during the rest trials. If the HFB response of an electrode was significantly correlated to the pronunciation of one or more of the sounds ($\alpha=0.05$, false discovery rate corrected for the number of retained electrodes), it was included in the subsequent analysis.

Determining Peak Timing

It is known that the HFB response peak related to the pronunciation of a sound does not always occur exactly at voice onset and can be different for different locations in the sensorimotor cortex (Crone et al. 1998; Coudé et al. 2011; Hermes et al. 2012; Bouchard et al. 2013). Yet, for the current study, we chose to analyze (and classify, as described below) the same point in time for all electrodes, and initially we analyzed the neural activity pattern for each sound only at one point in time (for other points in time see below). That is, one time-point for the first sound and one time-point for the second sound. These were the time points where the HFB amplitude was the highest in the average signal of all included electrodes. This was done because for fast decoding of sounds in (future) real-time BCI applications, each incoming time point (or small window) needs to be classified and the easiest way to do this is to compare each incoming time point's brain activity pattern to a single time window for each of the to be classified states. Thus, for the current analysis we determined the time point of maximal HFB activity related to the pronunciation of the first sound and the second sound, in the average signal of all included electrodes. This was done by epoching, for all included electrodes, the active trials twice: once aligned to the first sound (/u/ or /a/) onset, and once to the transition to the /i/ sound. We then took the average activity over all electrodes and trials and determined the timing of the mean HFB response peak closest to the sound onset and the mean HFB peak closest to the sound transition, respectively. Since we know from previous research (Salari et al. 2018b) that repeated isolated sounds are accompanied by a peak neural response on or just after voice onset, we assume that the same should be true for voice transitions.

Sensorimotor Cortex Activity Pattern Analysis

We determined for all trials (aligned to the sound transition) and per electrode, the HFB amplitude related to the /i/ pronunciation, by computing per included electrode and for each trial, the average HFB amplitude in a window of 0.3 seconds centered around the previously determined HFB response peak timing. Then, we determined whether there was a significant difference in the HFB spatial patterns between the three different conditions (/i/ preceded by silence, /i/ preceded by /u/ or /i/ preceded by /a/). We reasoned that if the HFB patterns for the different /i/ conditions would be significantly different, it should be feasible to classify which (if any) sound came before the /i/, based on the HFB pattern related to the /i/ sound. If the HFB pattern for /i/ would be unaffected by prior pronunciations, however, classification results of the three /i/ conditions should be at chance level. Therefore, we classified for each /i/ pronunciation the preceding sound (or silence), based on the HFB amplitude pattern associated with the /i/ sound, using a leave-one-trial out, template

matching classification procedure. We calculated for each trial the correlation with three templates (one for each /i/ condition) and assigned each trial to the template (class) with the highest correlation. We chose correlation as a distance measure since it has previously been used to successfully classify different sounds and hand gestures based on sensorimotor cortex neural activity (Bleichner et al. 2016; Branco et al. 2017; Ramsey et al. 2018). Templates of the HFB spatial pattern were made for each /i/ condition which was done by calculating the average (per electrode) HFB activity of all trials per class (/i/ condition), excluding the test trial. The averaged activities of all electrodes formed the spatial pattern template. Note that the template of the test trial's condition was based on one trial less than the other templates since for that condition, by consequence, there was one trial less available for training. However, we chose this leave-one-trial out classification procedure to maximize the number of training trials, given the relatively small number of trials in total. To determine chance level and classification accuracy significance, we performed a Monte Carlo procedure in which we shuffled the labels and repeated the classification a thousand times. Chance level is the mean of the Monte Carlo accuracies and significance level (or p-value) was determined by counting the number of Monte Carlo accuracies higher than the actual accuracy, divided by the number of iterations (1000).

Anatomical Localization of Differences

We investigated which electrodes showed differences in HFB amplitude between the three /i/ conditions, to see specifically which sensorimotor cortex areas show differences in HFB amplitude and whether or not there was an anatomical organization in the effects. For that, we used, per electrode, an ANOVA (alpha = 0.05, FDR corrected), in which the dependent variable was the HFB amplitude associated with the pronunciation of an /i/ sound, and the independent variable was the experimental condition.

Classification of Different Sounds

Next, we investigated if the /i/ sound patterns could be distinguished reliably from the patterns of the other two sounds (i.e. /a/ and /u/). In other words, we investigated whether or not the (different) patterns for the /i/ sound contained a unique signature and were more similar to each other than to the patterns of other sounds. To that purpose, we calculated the HFB amplitude patterns for the other two sounds in a similar way as explained for the /i/ sound, but now with trials aligned to the voice onset instead of sound transition. Subsequently we performed a leave-one-trial-out, template matching classification of the three different sounds (/i/, /a/, and /u/) based on the HFB amplitude patterns. If the (different) patterns for the three /i/ conditions contained a signature that is distinct from the /u/ and /a/ sound, all /i/ productions (no matter what sound came before it) should be classifiable using a template based on the data of only one of the /i/-combinations (for instance only on the isolated /i/ productions). However, this may not necessarily be the most optimal method for classification since the variation in patterns for the same sound is then not taken into account. Therefore, an individual template for each /i/ combination might be more optimal. Considering the number of sound combinations in natural speech, a separate template for each sound combination may, however, not be practical for a BCI. It is therefore interesting to see if a single template,

derived from the average of sound combinations, can be used for accurate classification, as this takes into account the variations in patterns, yet only uses one template per sound. To investigate these questions, we used three different methods to create the template(s) for the /i/ sound patterns. In the first method (the ‘Isolated /i/ method’), we used only the trials of the isolated /i/ (i.e. the one preceded by silence) to create the /i/ template. For the creation of the other two templates (for /u/ and /a/), the data during the production of the /u/ and the /a/ were used. Note that in this case, even though the template for the /i/ sound was based only on isolated /i/ productions, all /i/-combinations (/i/ preceded by silence, by /u/ or by /a/) were included in the classification of the three different sound productions (/i/, /u/ and /a/), meaning that the number of /i/ productions was three times larger than that of each of the other sounds. A correction for this difference in number of trials was applied, as described below. In the second method (the ‘Multiple /i/ method’), we used five different templates in total: three different templates for the /i/ sound (one for each of the conditions) and one for each of the other two sounds. In this case, if the test trial had the highest correlation with one of the three /i/ templates, the test trial was classified as belonging to the /i/ sound. In the third method (the ‘Mean /i/ method’), we took the average /i/ sound pattern across all three /i/ conditions as the /i/ template. By using these different methods for creating the /i/ template, we could determine if and how multiple activity patterns for one sound need to be incorporated in a BCI.

We compared the classification performance using the three different methods for creating the /i/ template with three different T-tests (comparing method 1 vs 2, 1 vs 3 and 2 vs 3, $\alpha=0.05$, FDR corrected). Since the /i/ sound has three times the number of productions as the other two sounds, we corrected the classification accuracy for the different number of productions per sound. This was done by taking first the accuracy of every class (the number of correctly classified trials of one class divided by the total number of trials within that class) and then taking the average of that as the overall accuracy (as opposed to taking the total number of correctly classified trials divided by the total number of trials). We determined the chance level with a Monte Carlo procedure, shuffling the labels 1000 times.

Finally, to see if all patterns (the three different /i/ patterns and the patterns for the /u/ and /a/) could be differentiated from each other, we also performed a classification of all five patterns, see supplementary material Figure S1.

Classification of the Different Conditions Over Time

In the above analyses, sound templates were based on the moment of peak neural activity, which is, likely, closely related to movements (articulator transitions), rather than to steady state pronunciation. To get some insights, however, in how long after the transition the different /i/ condition neural activity patterns stayed different we performed a classification over time between the three conditions while at the same time moving the templates in time (from 2 seconds before transition cue to 2 seconds after, in steps of 0.05 seconds). This analysis enabled us to see if the within-sound classification (of the three /i/ conditions) was based mostly on the transition phase or was also possible at later points in time (e.g. the steady state phase). To be able to compare results over trials, we epoched all trials, and corrected differences in voice onset, transition and offset in the brain recordings by down sampling and interpolation, similar as in Salari et al. (2018a), effectively aligning all voice onsets, transitions and offsets to their respective mean. The

significance level for classification was set to the 95% confidence interval on the binomial distribution above chance level.

Results

Task Performance and Electrode Selection

Participants performed the task adequately. Only one trial for one subject (C) was excluded due to a missed response. On average, subjects produced the /u/ for 0.64 (SD=0.13) seconds, the /a/ for 0.65 (SD=0.13) seconds and the /i/ for 0.67 (SD=0.18) seconds. Three subjects showed a significant difference between the durations of the different sounds (A: $F(2,147)=8.26$, $p<0.001$, C: $F(2,145)=27.65$, $p<0.001$ & E: $F(2,147)=29.85$, $p<0.001$). For the three different /i/ conditions, subjects produced the /i/ preceded with silence for 0.75 (SD=0.19) seconds, preceded by /u/ for 0.62 (SD=0.13) seconds and preceded by /a/ for 0.64 (SD=0.13) seconds. Two subjects showed a significant difference between the speech durations of the /i/ conditions (B: $F(2,42)=29.49$, $p<0.001$ & E: $F(2,87)=14.88$, $p<0.001$).

For subjects A-E respectively, 72% (18/25), 61% (11/18), 47% (17/36), 48% (55/114) and 89% (108/122) of the electrodes over the sensorimotor cortex showed a significant correlation with the task. Only these were used for subsequent analyses.

Classification of the Different /i/ Combinations

Figure 1 shows that for most subjects, there are clearly two peaks of activity (one per sound) and it also suggests that there are on average (across electrodes) differences between conditions. These differences are not equal over subjects though. Classification of the HFB patterns of the three /i/ sound conditions showed that, indeed, for all participants there were significant differences between these HFB patterns (Table 1 and Figure 1), indicating that prior sound production influences the brain activity pattern for the /i/ sound.

Investigation of whether or not differences between conditions occurred in all electrodes revealed that, of all electrodes, 22% (4/18), 27% (3/11), 29% (5/17), 45% (25/55) and 13% (14/108) showed a significant difference in HFB amplitudes between the three different /i/ production conditions (Figure 2). Notably, these electrodes appeared to be localized in clusters, which is especially clear in the HD-grid participants.

Classification of /a/, /u/ and /i/ Sound Production

Investigating the differences of the /i/ sound pattern(s) with that of the other sounds revealed that all subjects show significant classification accuracies for all three template methods (Figure 3 and 4 and Table 2). In other words, despite the fact that /i/ production is associated with different HFB patterns that depend on prior utterance, the /i/ sound can be distinguished reliably from the /a/ and the /u/, indicating that the different neural patterns for the /i/ sound are more similar to each other than to the other sounds. However, for the ‘Isolated /i/ method’, significantly lower classification accuracies were obtained ($M=63.05\%$, $SD=11.17\%$) than for the ‘Multiple /i/ method’ ($M=77.82\%$,

SD=13.19%, T= -5.02, p=0.01), or for the ‘Mean /i/ method’ (M=76.64%, SD=11.41%, T= -6.33, p<0.001). There was no significant difference in classification accuracy between the latter two methods (Figure 4).

Table 1. Classification of the different /i/ patterns. For all subjects (A-E), the classification accuracy, chance level and p-value (as determined by a Monte Carlo procedure) are shown, for the classification of the different /i/ productions. Significant classification of these patterns indicates that there are overall differences between HFB patterns associated with /i/ sound productions, indicating these patterns depend on the previous sound.

Participant	Accuracy (%)	Chance level (%)	p
A	48.89	33.11	0.018
B	64.44	33.12	0.003
C	58.20	33.14	0.001
D	83.33	33.08	<0.001
E	75.56	33.23	<0.001

Table 2. Statistics for the different sound patterns classification, using three different /i/ template methods. For all subjects (A-E) the classification accuracy is indicated for the classifications of the different sounds per /i/ template method. Also, the average accuracy (over subjects) is shown. Chance level, as determined by a Monte Carlo procedure is shown between brackets. Asterisks indicate a classification accuracy significantly above chance (*p<0.01, **p<0.001).

Participant	Isolated /i/	Multiple /i/	Mean /i/
A	50.74% (33.14%)*	61.85% (33.31%)**	64.44% (33.38%)**
B	60.00% (33.05%)*	74.81% (33.45%)**	71.11% (33.22%)**
C	63.74% (33.19%)**	71.34% (33.40%)**	72.47% (34.53%)**
D	81.11% (33.15%)**	96.30% (33.23%)**	94.07% (33.72%)**
E	59.63% (33.29%)**	84.81% (33.15%)**	81.11% (32.97%)**
Mean	63.05% (33.16%)	77.82% (33.31%)	76.64% (33.57%)

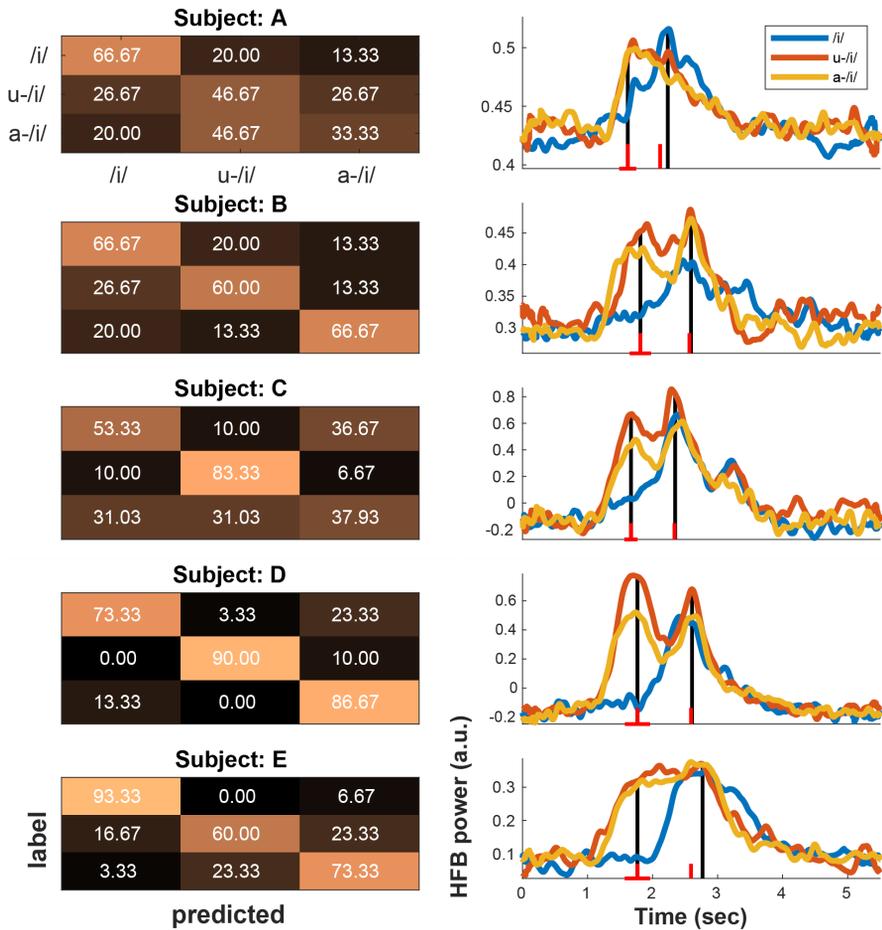


Figure 1. Confusion matrices and average response for the different /i/ productions. For each participant (A-E) the confusion matrix for classification of the different /i/ HFB patterns is shown on the left. The numbers indicate the percentage (%) of trials within the class (indicated by the rows) that were classified as the class indicated by the columns. On the right, the average HFB response (over electrodes and trials) is shown for each participant and each condition, with time on the x-axis and the HFB-power (in arbitrary units) on the y-axis. Voice onset and voice transition are shown by vertical red lines and the corresponding response peak times are shown by vertical black lines. Trials were aligned to voice transition and the voice onset standard deviation is indicated with a red horizontal bar. Note that since all trials are aligned to voice transition, there is no standard deviation for this. The sound before the /i/ is indicated without slashes to highlight the fact that classification specifically focused on the /i/ sound here.

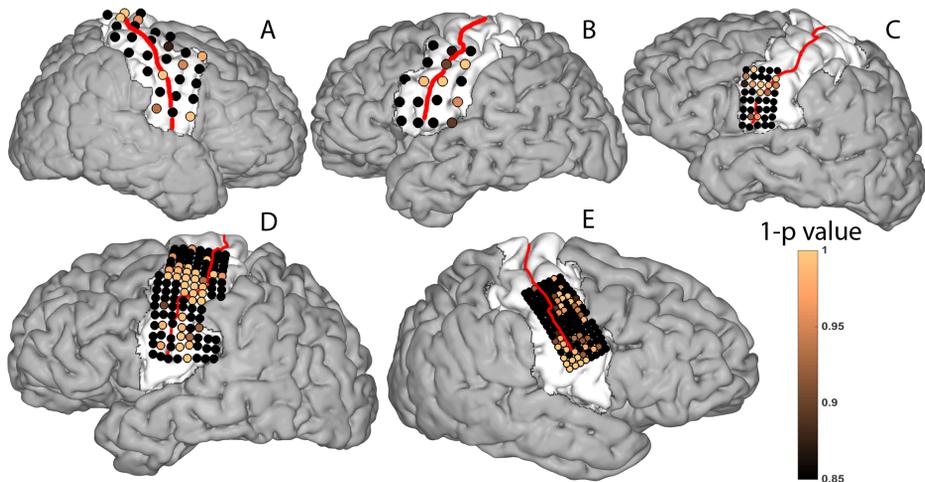


Figure 2. Sensorimotor cortex electrodes that show differences in HFB amplitude between the different /i/ sound conditions. The brain surface rendering of each subject (A-E) is shown. The sensorimotor cortex is highlighted in white and the central sulcus in red. For each electrode, an ANOVA was performed to test for differences between the HFB amplitudes during the different /i/ productions and the inverted p-values (1-p) thereof are shown here by the color shading of each electrode. Lighter colors represent statistically stronger differences between different /i/ sound conditions.

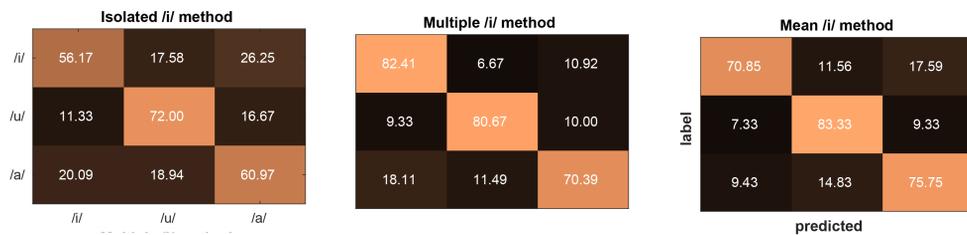


Figure 3. Average confusion matrices for classification of the three different sounds using the different /i/ template methods. The average (over subjects) confusion matrices are shown for each of the three different methods used to create /i/ templates. The numbers indicate the percentage (%) of trials within the class that is indicated by the rows that were classified as the class indicated by the columns.

Classification at Different Points in Time

To see if the classification of the three /i/ conditions was based mostly on the transition phase or was also possible at later points in time (e.g. the steady state phase) we did a classification over time between the three conditions using different time points to generate the templates (as opposed to using the peak neural activity). Results showed that, using this approach, the classification

accuracy of the three /i/ conditions was highest at the neural peak activity (which was usually very close to the voice transition) and that it dropped rapidly after the peak, suggesting that the difference of neural activity between the /i/ conditions is mostly the result of differences in the movement of the articulators toward the sound (Figure 5). However, the different /i/ conditions stayed significantly separable between 0.28 - 1.01 seconds after the transition.

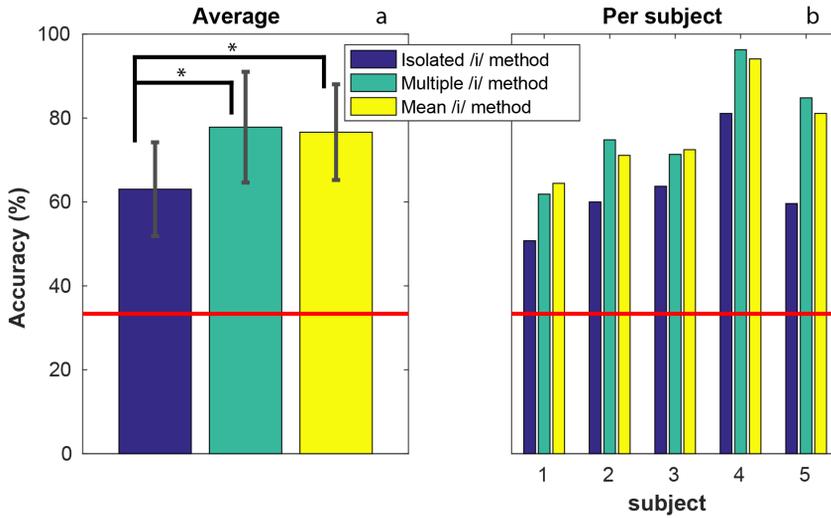


Figure 4. Average and individual classification accuracies for classification of the different sounds using the three different /i/ template methods. In panel (a) the average (over subjects) classification accuracies are shown for each of the three different /i/ template methods. In panel (b) the same is shown for each individual separately. In panel (a), an asterisk indicates a significant difference in accuracy between methods. The horizontal red line indicates the average chance level (over subjects and methods).

Discussion

In this study, using a simple speech task we confirmed that sensorimotor neural signals related to sound production are affected by previous sound productions. Electroencephalography measurements in five epilepsy patients revealed that the HFB amplitude pattern in the sensorimotor cortex associated with production of a certain sound may differ, depending on the presence and nature of a preceding sound production. These data indicate that production of the same sound is associated with multiple HFB amplitude patterns in the (ventral) sensorimotor cortex, likely due to differences in the movements towards a sound, which is of direct relevance for the development of speech BCI. This phenomenon was not only observed in the averaged signal of all electrodes, but also at single electrode level. The anatomical localization of electrodes that showed a significant difference in HFB amplitude during the different productions of the /i/ sound seemed to be somewhat clustered for most subjects, but given the limited number of participants, it cannot be concluded whether or not there is a typical anatomical distribution for the effects we found. Importantly, the differences in HFB amplitude between the three different /i/ conditions were most apparent just after voice

transition (close to the moment of peak neural activity) and classification accuracy decreased rapidly after the transition. This suggests that the difference in neural activity between the /i/ conditions is mostly due to the articulator movements toward the sound, and that, once the articulators are in place, pronunciation of the /i/ sound is less affected by prior utterances. For most subjects, however, the different /i/ conditions stayed significantly distinguishable for 0.28 – 1.01 seconds after the transition, suggesting that there might be some differences in the neural activity that are not only due to the transition.

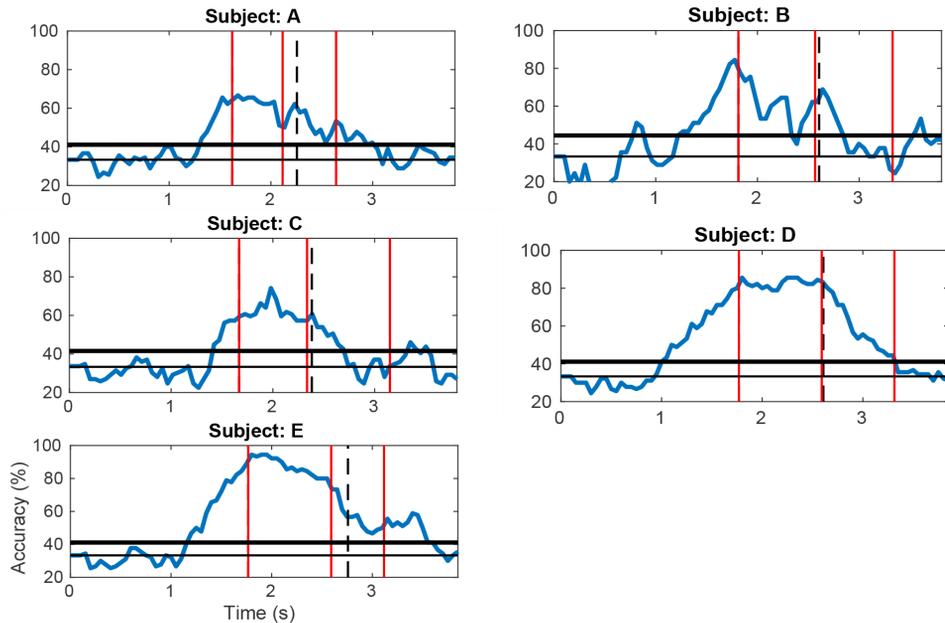


Figure 5. The classification accuracy between the three conditions over time, using templates generated at moving time points. The classification accuracy (y-axis) is shown at each time point (x-axis) between the three different conditions (/i/, /u/-/i/ & /a/-/i/), for each subject. Time point 0 is 2 seconds before transition cue onset. Templates were based on the same point in time as the time point of classification (rather than only at the peak neural activity, as used earlier). Voice onset, transition and offset are indicated by vertical red lines. The point at which originally, the /i/ templates were based (the peak in neural activity) is indicated by a dashed black line. Note, that for most participants the most separable moment was around the peak of neural activity. For subject E, the peak of neural activity was rather wide (see figure 1) which may explain the time difference between the accuracy peak and the neural response peak. Chance level (33.33%) is indicated by a thin horizontal line and the thick horizontal line indicates the accuracy level at which classification is significantly above chance level. Note that, at the onset of a trial, classification should be at chance level, since at this moment all three templates were based on rest and should not be able to predict the future, upcoming, pronunciation. Between voice onset and voice transition, the classification is between rest, the /u/ and the /a/ sound. Between voice transition and voice offset the classification is between the three different /i/ conditions (the isolated /i/, the /i/ with an /u/ before and the /i/ with an /a/ before), basically predicting if and what sound was said in the past (just before the /i/ sound).

Interestingly, despite the different response patterns associated with /i/ pronunciation, we found that the different sounds (/i/, /u/, /a/) can be classified accurately from one another, irrespective of whether the /i/ was preceded by another sound or not. This means that the representation of /i/ production in the sensorimotor cortex contains a distinct signature that can be distinguished from the representation of /a/ and /u/ production, and would seem to be independent of preceding sound production. However, we also show that classification accuracies for the different sounds increase significantly when the multiple HFB amplitude patterns for the /i/ sound are taken into account. These results suggest that, for the purpose of speech BCI, the training of a classifier benefits from taking into account the differences in sound representation that are caused by preceding speech production, since this will lead to higher classification accuracies.

Neural Underpinnings of Multiple HFB Amplitude Patterns

The different HFB amplitude patterns related to the production of a single vowel may be attributed to coarticulation. Coarticulation can occur in two directions: the pronunciation of a previous sound can influence the articulator position of an upcoming sound (carryover) or vice versa (anticipatory) (Rietveld and van Heuven 2001; Davis 2015). Literature from linguistic experiments have shown for instance that the tongue position during the pronunciation of the /k/ in the word ‘seek’ is higher than in the word ‘arc’, since the /i/ is a fronted vowel (with the tongue front high) and the /a/ has a low tongue position and the tongue position during the vowel extends to the upcoming sound(s) (example taken from Davis 2015). An example of anticipatory coarticulation can be seen during the pronunciation of the words ‘ski’ and ‘cool’ in which the /k/ is more fronted (tongue front is high) in ‘ski’ due to an anticipation of the fronted vowel /i/ and less fronted in ‘cool’, in anticipation of the /u/ which has a more backward tongue position (Davis 2015). Recent investigations have shown that articulator position differences due to coarticulation can lead to different levels of activity for the same sound in some parts of the sensorimotor cortex (Bouchard and Chang 2014; Chartier et al. 2018). Here we extend those results by showing where these differences may be anatomically localized for one sound and how these effects can influence decoding accuracy for speech BCI. In addition, we show that, even though coarticulation effects are present in the sensorimotor cortex, the different patterns for one particular sound have something in common that distinguishes them from the patterns of other sounds.

Another explanation for the current results could be the fact that the movements of the articulators to a particular sound differ depending on the prior sound. For instance, it is known for the Dutch language that the tongue position when producing an /a/ is mostly low (somewhat similar as the rest position), while for the /u/ sound the back of the tongue moves up the most and for the /i/ sound the front of the tongue moves up the most (Booij 1999). This means that the movement from an /a/ to an /i/ sound may be more similar to producing solely the /i/ sound, than to that of producing an /u/ sound before the /i/ sound. However, even though these movement differences are likely to be quite similar across subjects, there was no convincing evidence in our study for a consistently higher similarity of some of the brain activity patterns (e.g. those of the isolated /i/ and the /i/ with the /a/ before it), compared to others (e.g. the /i/ with the /u/ before it). Differences in electrode positions and coverage over subjects, as well as the sparse spatial sampling of ECoG, may have influenced the current results, however, since only when similar parts of the cortex are

recorded it is possible to compare patterns over subjects.

Implications for BCI

The current data contribute to a better understanding of the activity properties of the sensorimotor cortex and this may be of interest for the development of BCIs that aim to decode speech sounds based on sensorimotor cortex activity. We show that multiple sensorimotor cortex HFB amplitude patterns exist for pronunciation of a single sound, which are likely related to the movements toward that sound, and that classification of different sounds can benefit from taking these different patterns into consideration. This does not mean however, that one should employ separate templates for all combinations of sounds. Our results indicate that an average template of the different patterns will suffice for accurate classification. Previous studies that have trained a classifier for phoneme classification not based on isolated phonemes but on phonemes within a word or sentence (e.g. Mugler et al. 2014; Herff et al. 2015), implicitly took some context for sound production in consideration. Our results suggest that explicit and careful consideration of different contexts is essential for accurate sound classification.

Limitations and Future Work

A higher number of participants and a more extensive electrode coverage would have strengthened the current results. In addition, in the current experiment we tested only a limited number of sound combinations. Further research is needed to investigate the effects of prior sound production on other phonemes than the /i/ and the influence of this on BCI classification accuracy. Another limitation is that, although we tried to stay as constant as possible with respect to marking the transitions, finding the exact transition onset is not always clear and might therefore have caused some variability in transition onset marking. Finally, as mentioned above, coarticulation effects can also occur in an anticipatory fashion, meaning that the sounds that follow a particular sound influence the articulator position of the prior sound. More research on the consequences of such an anticipatory coarticulation on BCI classification seems warranted.

Conclusion

We demonstrate that the production of one sound is accompanied with multiple sensorimotor cortex activity patterns. This is of importance for BCIs that aim to decode speech from the sensorimotor cortex since we show that multiple movements related to the same sound can be classified into a single vowel sound class. Classification is, however, further improved when the multiple patterns for one sound are taken into account by the classifier (either by training a classifier for the separate patterns of the same sound or by training on the averaging of the multiple patterns for one sound).

Acknowledgements

Our thanks go out to the participants for their contribution and the clinical neurophysiology department staff and the neurosurgeons for their collaboration and cooperation.

References

- American Congress of Rehabilitation Medicine (1995) Recommendations for use of uniform nomenclature pertinent to patients with severe alterations in consciousness. *Arch Phys Med Rehabil* 76:205–209. doi: 10.1016/S0003-9993(95)80031-X
- Bleichner MG, Freudenburg ZV, Jansma JM, et al (2016) Give me a sign: decoding four complex hand gestures based on high-density ECoG. *Brain Struct Funct* 221:203–216. doi: 10.1007/s00429-014-0902-x
- Boersma P (2002) Praat, a system for doing phonetics by computer. *Glott Int* 5.
- Booij G (1999) *The Phonology of Dutch*. Clarendon Press
- Bouchard KE, Chang EF (2014) Control of Spoken Vowel Acoustics and the Influence of Phonetic Context in Human Speech Sensorimotor Cortex. *J Neurosci* 34:12662–12677. doi: 10.1523/JNEUROSCI.1219-14.2014
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–332. doi: 10.1038/nature11911
- Branco MP, Freudenburg ZV, Aarnoutse EJ, et al (2018a) Optimization of sampling rate and smoothing improves classification of high frequency power in electrocorticographic brain signals. *Biomed Phys Eng Express* 4:045012. doi: 10.1088/2057-1976/aac3ac
- Branco MP, Freudenburg ZV, Aarnoutse EJ, et al (2017) Decoding hand gestures from primary somatosensory cortex using high-density ECoG. *NeuroImage* 147:130–142. doi: 10.1016/j.neuroimage.2016.12.004
- Branco MP, Gaglianesi A, Glen DR, et al (2018b) ALICE: A tool for automatic localization of intra-cranial electrodes for clinical and high-density grids. *J Neurosci Methods* 301:43–51. doi: 10.1016/j.jneumeth.2017.10.022
- Brumberg JS, Wright EJ, Andreasen DS, et al (2011) Classification of Intended Phoneme Production from Chronic Intracortical Microelectrode Recordings in Speech-Motor Cortex. *Front Neurosci* 5. doi: 10.3389/fnins.2011.00065
- Bruns A (2004) Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? *J Neurosci Methods* 137:321–332. doi: 10.1016/j.jneumeth.2004.03.002
- Chartier J, Anumanchipalli GK, Johnson K, Chang EF (2018) Encoding of Articulatory Kinematic Trajectories in Human Speech Sensorimotor Cortex. *Neuron*. doi: 10.1016/j.neuron.2018.04.031
- Coudé G, Ferrari PF, Rodà F, et al (2011) Neurons Controlling Voluntary Vocalization in the Macaque Ventral Premotor Cortex. *PLoS ONE* 6:e26822. doi: 10.1371/journal.pone.0026822
- Crone NE, Miglioretti DL, Gordon B, Lesser RP (1998) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* 121:2301–2315. doi: 10.1093/brain/121.12.2301
- Davis J (ed) (2015) *Phonetics: fundamentals, potential applications and role in communicative disorders*. Nova Publishers, New York
- Herff C, Heger D, de Pestiers A, et al (2015) Brain-to-text: decoding spoken phrases from phone representations in the brain. *Neural Technol* 9:217. doi: 10.3389/fnins.2015.00217
- Hermes D, Miller KJ, Noordmans HJ, et al (2010) Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J Neurosci Methods* 185:293–298. doi: 10.1016/j.jneumeth.2009.10.005
- Hermes D, Siero JCW, Aarnoutse EJ, et al (2012) Dissociation between Neuronal Activity in Sensorimotor Cortex and Hand Movement Revealed as a Function of Movement Rate. *J Neurosci* 32:9736–9744. doi: 10.1523/JNEUROSCI.0357-12.2012
- Hermes DJ (1990) Vowel-onset detection. *J Acoust Soc Am* 87:866–873. doi: 10.1121/1.398896
- Kellis S, Sorensen L, Darvas F, et al (2016) Multi-scale analysis of neural activity in humans: Implications for micro-scale electrocorticography. *Clin Neurophysiol* 127:591–601. doi: 10.1016/j.clinph.2015.06.002
- Lee KF (1990) Context-independent phonetic hidden Markov models for speaker-independent continuous speech recognition. *IEEE Trans Acoust Speech Signal Process* 38:599–609. doi: 10.1109/29.52701
- Manning JR, Jacobs J, Fried I, Kahana MJ (2009) Broadband shifts in LFP power spectra are correlated with single-neuron spiking in humans. *J Neurosci Off J Soc Neurosci* 29:13613. doi: 10.1523/JNEUROSCI.2041-09.2009
- Miller KJ, Leuthardt EC, Schalk G, et al (2007) Spectral Changes in Cortical Surface Potentials during Motor Movement. *J Neurosci* 27:2424–2432. doi: 10.1523/JNEUROSCI.3886-06.2007
- Mugler EM, Patton JL, Flint RD, et al (2014) Direct classification of all American English phonemes using signals from functional speech motor cortex. *J Neural Eng* 11:035015. doi: 10.1088/1741-2560/11/3/035015
- Penfield W, Boldrey E (1937) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain J Neurol* 60:389–443. doi: 10.1093/brain/60.4.389
- Posner JB, Plum F, Saper CB, Schiff N (2007) *Plum and Posner’s Diagnosis of Stupor and Coma*. Oxford University Press, USA

- Ramsey NF, Salari E, Aarnoutse EJ, et al (2018) Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage*. doi: 10.1016/j.neuroimage.2017.10.011
- Ray S, Maunsell JHR (2011) Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLOS Biol* 9:e1000610. doi: 10.1371/journal.pbio.1000610
- Rietveld A, van Heuven V (2001) *Algemene Fonetiek*, 2nd edn. Coutinho, Bussum
- Salari E, Freudenburg ZV, Vansteensel MJ, Ramsey NF (2018a) Spatial-Temporal Dynamics of the Sensorimotor Cortex: Sustained and Transient Activity. *IEEE Trans Neural Syst Rehabil Eng* 26:1084–1092. doi: 10.1109/TNSRE.2018.2821058
- Salari E, Freudenburg ZV, Vansteensel MJ, Ramsey NF (2018b) Repeated Vowel Production Affects Features Of Neural Activity In Sensorimotor Cortex. *Brain Topogr*
- Schultz T, Wand M (2010) Modeling coarticulation in EMG-based continuous speech recognition. *Speech Commun* 52:341–353. doi: 10.1016/j.specom.2009.12.002
- Siero JC, Hermes D, Hoogduin H, et al (2014) BOLD matches neuronal activity at the mm scale: A combined 7 T fMRI and ECoG study in human sensorimotor cortex. *NeuroImage* 101:177–184. doi: 10.1016/j.neuroimage.2014.07.002
- Smith E, Delargy M (2005) Locked-in syndrome. *BMJ* 330:406–409. doi: 10.1136/bmj.330.7488.406
- Towle VL, Yoon H-A, Castelle M, et al (2008) ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain* 131:2013–2027. doi: 10.1093/brain/awn147
- Vansteensel MJ, Pels EGM, Bleichner MG, et al (2016) Fully Implanted Brain–Computer Interface in a Locked-In Patient with ALS. *N Engl J Med*. doi: 10.1056/NEJMoa1608085
- Wolpaw JR, Birbaumer N, McFarland DJ, et al (2002) Brain–computer interfaces for communication and control. *Clin Neurophysiol* 113:767–791. doi: 10.1016/S1388-2457(02)00057-3

Appendix

All patterns classification

	<i>/i/</i>	<i>u-/i/</i>	<i>a-/i/</i>	<i>/u/</i>	<i>/a/</i>
<i>/i/</i>	50.00	7.33	15.33	6.67	20.67
<i>u-/i/</i>	10.67	58.00	15.33	9.33	6.67
<i>a-/i/</i>	13.49	18.16	58.92	4.02	5.40
<i>/u/</i>	4.00	4.00	1.33	80.67	10.00
<i>/a/</i>	12.05	4.02	2.05	11.49	70.39
	<i>/i/</i>	<i>u-/i/</i>	<i>a-/i/</i>	<i>/u/</i>	<i>/a/</i>

predicted

Figure S1. Average confusion matrix for the classification of the sensorimotor cortex HFB amplitude patterns associated with all conditions. The average (over subjects) confusion matrix is shown for the classification of sensorimotor cortex HFB amplitude patterns of all conditions (i.e. three different */i/* conditions, */a/* and */u/*). The sounds preceding the */i/* sound are indicated without slashes. The numbers indicate the percentage (%) of trials within the class that is indicated by the rows that were classified as the class indicated by the columns. For all subjects, we could significantly classify all HFB amplitude patterns (*/a/*, */u/* and 3 different */i/* conditions) from each other above chance level, $p < 0.001$. The mean accuracy (over subjects) was 63.60% ($SD = 16.26\%$) and the chance level was 19.96% ($SD = 0.25$).





Chapter 4

*“I am you and you are mine, love makes nonsense of space and time will disappear.
Love and logic keep us clear. Reason is on our side, love.”*

Chapter 4

Classification of Articulator Movements and Movement Direction from Sensorimotor Cortex Activity

E. Salari, Z.V. Freudenburg, M.P. Branco, E.J. Aarnoutse, M.J. Vansteensel, N.F. Ramsey

Submitted Article

Abstract

Background: During human communication through speech, a lot of muscles are involved in the coordination of articulator movements. The sensorimotor cortex is important in the control of articulator movements and has a topographical representation for, for instance, the lips, tongue, jaw and larynx. However, the spatial scale at which these articulator movements are represented on the cortical surface remains unclear, as is the scale of within-articulator movements. **Methods:** Therefore, we recorded with subdural electrocorticography (ECoG) electrode grids, the neural activity related to movements of the lips, jaw, tongue and larynx in 4 epilepsy patients who were implanted with ECoG grids. We classified which movements participants made based on the sensorimotor cortex activity patterns, using a template matching procedure. The same was done for different movement directions of one and the same articulator, the tongue. **Results:** Both experiments showed high accuracy scores (on average 92% for different articulators and 83% for different tongue movements). Furthermore, we show that only a small part of the sensorimotor cortex is needed for classification (ca. 1 cm²). **Discussion:** Our results contribute to a better understanding of the functional organization of the sensorimotor cortex and are of interest for brain-computer interface systems which decode neural activity related to (attempted) movements, preferably from a contained cortical area. We show that recordings from small parts of the sensorimotor cortex can be used for distinguishing different articulator movements and different movement directions of the tongue.

Keywords: Sensorimotor Cortex, Movement, Articulators, ECoG, Tongue

Introduction

Even though most people can speak without much effort, the control of articulator and other movements during speech is very precise and complex, and has been suggested to involve more than one hundred muscles (Levelt 1993, p. 413; Meister et al. 2007; Guenther and Hickok 2015, p. 161). Research investigating the neural basis of articulator movements has found that the sensorimotor cortex (SMC), especially the ventral part thereof (vSMC), plays a key role in the control of articulator movements (although also the SMC hand area has been suggested to be involved; Saarinen et al. 2006). An early mapping study of Penfield and Boldrey (1937), showed that electrical cortical stimulation of the SMC may result in vocalization or in motor and sensory responses of the tongue, lips and jaw. More recently, electrocorticographic (ECoG) recordings suggested that the articulators are topographically organized in the vSMC, with a ventral to dorsal distribution of respectively the larynx, tongue, jaw, lips and larynx again (Bouchard et al. 2013). Although investigations using fMRI blood oxygenation level dependent (BOLD) signals have indicated overlap between the areas involved in different articulator movements, these studies have also largely confirmed the ventral-dorsal topographic organization of the vSMC (Grabski et al. 2012), and showed that the sensorimotor cortex neural activity patterns for different articulators can be distinguished (classified) from each other (accuracy ca. 90%; Bleichner et al. 2015). Indeed, since the production of different words and sounds is accompanied by differences in articulator movements (Mermelstein 1973; Levelt 1993, pp. 422–434; Booij 1999; Rietveld and van Heuven 2001), and because words and sounds can be distinguished from each other based on SMC activity (e.g. Kellis et al. 2010; Brumberg et al. 2011; Mugler et al. 2014; Herff et al. 2015; Ramsey et al. 2018) it can be speculated that the engagement of different articulator movements is differently represented in the SMC. However, details about subareas of the SMC that contain information about movements of different articulators, such as where these subareas are located or what size they have, remain scarce.

In the current study, we first investigate in more detail if movements that involve different articulator muscle groups can be distinguished from each other based on SMC activity. Furthermore, we investigate which SMC areas are most informative and at what spatial scale information of different articulator movements is represented. To assess cortical representation differences, we used classification as a measure of pattern distinguishability.

However, in addition to cortical pattern differences for different articulator movements, some studies have also suggested that during speech, different positions of the *same* articulator result in different vSMC activity patterns (Chartier et al. 2018, Salari et al 2018b). These studies have not looked, however, at articulator movements in isolation and at different directions of movements. Although previous single cell studies (Georgopoulos et al. 1982; Moran and Schwartz 1999) and fMRI studies (Toxopeus et al. 2011) have indicated that different directions of movements, mostly of the hand, have been found to be represented in the SMC, fewer studies have looked at a SMC direction representation of articulators with ECoG in humans. It remains to be elucidated, therefore, if different movement directions of the same articulator can be distinguished from each other based on sensorimotor cortex activity and if so, at what spatial scale this is possible.

Therefore, we also investigate if movements within one articulator muscle group are

differently represented in the SMC. We focused for this on the tongue and investigated if different directions of tongue movement can be distinguishing, thereby engaging different subgroups within the same articulator muscle group. We chose the tongue since this is an important articulator in differentiating sounds (Booij 1999; Rietveld and van Heuven 2001) and it can be predicted which sound somebody said based solely on information about tongue position (Wang et al. 2013). Furthermore, the tongue can easily be moved in different directions.

We analysed neural activity recordings of the sensorimotor cortex in 4 participants, while they moved four different articulators (lips, jaw, tongue and larynx), similar to those described by Bleichner et al. (2015), and investigated if they could be classified using electrophysiological surface measures. Furthermore, we analysed which SMC areas can be addressed for this and at what spatial scale this can be done. In addition, we analysed four different movement directions of the tongue (up, down, left and right), to investigate if these different movements could be classified from each other based on SMC neural activity. A template matching, leave-one-trial-out classification method was used to classify different articulator movements and different tongue movement directions based on spatial or spatio-temporal SMC activity patterns. High-density ECoG-electrodes were used to record brain activity. We focused on high-frequency band (HFB; 60-130 Hz) power changes in the SMC, since it has been shown to correlate well with articulator movements (Crone et al. 1998; Bouchard et al. 2013) and because it has been linked to neural firing (Manning et al. 2009; Miller et al. 2009; Ray and Maunsell 2011).

Methods

Participants

Four patients (A-D, age 19-41y; median 34.5y, 2 females) participated in this study, while they were treated in the University Medical Center Utrecht for epilepsy. Subdural ECoG electrodes were implanted sub-chronically (subject A, B and D) or were used during awake surgery (subject C) to record neural signals for clinical purposes. For research purposes, a non-clinical high-density (HD) electrode grid was placed (sub-chronically or temporarily during awake surgery) over a clinically non-relevant area (the sensorimotor cortex) with the consent of the participants. Electrodes were 1 or 1.17 mm in diameter with a 3- or 4-mm inter-electrode distance, see Table 1. For the current study, only signals recorded with the HD electrodes were used in the analyses. Data for subject C were recorded intraoperatively during the awake surgery. For the other subjects, the data was recorded at the patients' hospital bed during a week of clinical assessment.

This study was in accordance with the Declaration of Helsinki (2013) and was approved by the ethics committee of the University Medical Center Utrecht. Written informed consent was given by all subjects.

Task

The participants performed two tasks. First, we wanted to see if we could replicate earlier fMRI findings by Bleichner et al. (2015) showing that movements of different articulators can be classified from SMC neural activity. Therefore, in the first task (the 'Articulator Task'), subjects were asked to make four different articulator movements. The movements were (1) pursing the lips,

(2) clenching the teeth, (3) moving the tongue from left to right behind the teeth, (4) making a ‘mmm’ sound. These movements involve the lips, jaw, tongue and larynx, respectively, and were chosen to be similar to those used by Bleichner and colleagues (2015). The task was presented on a computer screen at a comfortable distance from the participants. During the intraoperative recordings, the task was presented on a tablet which was placed on a pole such that it was clearly visible to the participant. A trial started with a 1500 ms visual cue, being the Dutch word ‘lippen’ (lips), ‘tanden’ (teeth), ‘tong’ (tongue), or ‘mmm’. This visual cue instructed the subjects to start the movement and hold it for as long as the cue was visible. Movement trials were interleaved with trials in which the subject was instructed not to move, indicated by a ‘-’ symbol. Trials were followed by an inter-trial interval of 1500 ms (subjects A, B & D) or 2000 ms (subject C) during which a fixation cross was presented. Each movement was repeated 20 times in random order as were the rest trials.

In the second task (the ‘Tongue Task’), the subjects were instructed to make four different tongue movements inside the mouth. The movements were (1) up, (2) down, (3) left and (4) right from rest position. This task was similar to the first task except that the visual cues now were arrows pointing in each direction.

Some subjects performed some additional runs of the experiments, see Table 1. Subject C only performed one task (the Articulator Task) due to limited time in the intraoperative setting. Subjects A & B performed the tongue movement task also once with the tongue outside the mouth.

Table 1. Recording details. *In the top section, we show which subject did which task. In the bottom section, we indicate per subject the details of the electrode grids.*

Subject	A	B	C	D
Tasks performed				
Articulators run 1	X	X	X	X
Articulators run 2	X	X		
Tongue inside run 1	X	X		X
Tongue inside run 2		X		
Tongue outside	X	X		
Electrode diameter (mm)	1	1.17	1	1
Inter-electrode distance (mm)	4	4	3	3
Number of recorded electrodes	64	128	128	128

Data Acquisition & Preprocessing

The sampling frequency of the neural signal recordings was 512 Hz (subjects A; Micromed, Treviso, Italy) or 2000Hz (subjects B, C & D; Blackrock Microsystems LLC, Salt Lake City, USA). In total, we recorded from 64 electrodes for subject A and from 128 electrodes for subjects B-D. For all subjects except subject C, the electrode positions were identified from a post-operative computed tomography (CT) scan which were subsequently plotted over a 3D surface of the subjects pre-operative magnetic resonance imaging (MRI) scan (Hermes et al. 2010; Branco et al. 2018b).

For subject C, there was no CT scan as the data was recorded intraoperatively and electrode positions were determined using a correlation between an average HFB power grid pattern (during rest) and the underlying anatomical structure (e.g. the location of sulci or blood vessels) to estimate the electrode positions. This method has been validated for localization of HD electrodes (Branco et al. 2018c).

We used Matlab software (The Mathworks, Inc., Natick, MA, USA) for all data analysis. First, flat or noisy electrodes were removed from further analysis similar as in Salari et al (2018a). For the remaining electrodes, we applied a notch filter for the removal of line noise and harmonics thereof. For removal of any remaining artefacts, a common-average re-referencing was applied. Subsequently, we used a Gabor wavelet (Bruns 2004) to calculate the high frequency band (60-130 Hz) power per sample point for each included electrode. The Gabor wavelet was calculated with a full width half maximum (fwhm) of 4 wavelets per frequency in bins of 1 Hz for all frequencies between 60 and 130 Hz. The Gabor wavelet results were subsequently log transformed ($10 \cdot \log_{10}$) and averaged over frequencies to create the HFB power signal per electrode. Finally, we determined for each electrode if it was responsive to the task. To that purpose, we computed, per included electrode, the r^2 -value by correlating the mean activity levels during active periods (from cue until the end of a trial) and rest periods (rest trials) with the task design. This was done for each movement separately. We determined the significance level of each electrode by using a Monte Carlo distribution (shuffling active and rest labels using 10.000 permutations, Alpha=0.05, false discovery rate corrected). We calculated the r^2 -values to determine the activity patterns for each movement and for electrode selection during classification. Electrodes that did not show a significant HFB power response to any of the movements of a specific task, were removed from analysis of that task. For classification, we used a leave-one-trial-out procedure, so for electrode selection we determined the significant electrodes for each iteration separately based on the training set (not including the test trial).

Classification Procedures

For each of the included significant electrodes, we smoothed the signal with a moving average window of 0.5 seconds around each sample point, which has been shown previously to be an optimal setting for classification of movements from ECoG neural recordings (Branco et al. 2018a). Subsequently, we z-scored the smoothed signal and epoched the result in trials of 2 seconds, each starting at cue onset. Subsequently, we performed two types of leave one out template matching classification. First, we averaged the power signal of each trial for each electrode. Then, one trial was taken out of the data set and subsequently the spatial templates of the four conditions were created by averaging (per electrode) the values of all remaining trials. These templates were then correlated with the trial that was taken out and the template with the highest correlation to the trial was chosen in a winner-takes-all fashion. This was repeated until all trials were classified.

Second, as we wanted to know if timing information could lead to better classification scores, we repeated this classification method but without averaging the power over the trial period.

Anatomical Localization of Informative Electrodes

To investigate where exactly the most informative electrodes for articulator and tongue movement classification were located, we used a random search procedure in which we randomly chose a subset of electrodes to classify from and repeated this 5000 times. Subsequently, for each electrode we calculated what the average accuracy was if that electrode was part of the electrode set. These scores were subsequently z-scored to find the hotspots of electrodes that were the most informative. We investigated how the number of iterations for this procedure influenced the final pattern of most informative electrodes by determining the informative-electrode-pattern at each iteration and subsequently calculating the mean correlation of that pattern with all subsequent iterations patterns. This shows how much the pattern changes with each extra iteration. The less the pattern changes, the more reliable the pattern result is. We found that 5000 iterations gave a reliable result, see supplementary figure S1.

Anatomical Scale of Informative Areas

To know how much of the sensorimotor cortex needs to be sampled from, for accurate classification, we used a searchlight approach in which we started with a small grid size (only 1 electrode) and then increased the grid size, each time with one extra grid row and column of electrodes, until the original grid size was reached. This was done to find the size of the area that was needed for accurate classification, given that electrodes could not be anywhere, but were restricted to be closely located to each other. Per grid size and per location, we calculated the accuracy score (only for spatial classification). We determined per grid size what the maximal accuracy score was (best of all different sampled locations) to assess the minimally required grid size for accurate classification.

Topographical Organization of the Articulators

While previous research has indicated overlap between the areas that are active during movements of different articulators (Grabski et al. 2012), the articulators also show a topographical organization in the SMC (Grabski et al. 2012; Bouchard et al. 2013). We used a procedure described by Bruurmijn and colleagues (2017) to normalize the SMC of each subject into an isotropic coordinate system to investigate if we could also find a topographical organization of different movements. We created a group average activity pattern of the (sampled) SMC for each articulator movement and each tongue movement to see where each movement was mostly localized on average (over subjects). This was done by using the calculated electrodes' r^2 -values for each movement and plotting these within the normalized coordinate system. In addition, we made a winner-takes-all plot, indicating for each electrode, which movement had the highest r^2 -value. Only electrodes with a r^2 -value of more than 0.2 in at least one of the conditions were used for this analysis.

Results

Task Performance and Electrode Selection

In general, subjects did not report to have any difficulty performing the tasks, except for subject A who reported that the articulator task was a little fast. We did not exclude any trials based on performance, however, since performance of most of the movements had to be done with the mouth closed and were therefore not visible from the outside.

During the first run, 60.94% (39/64), 80.47% (103/128), 87.50% (112/128), and 84.38% (108/128), of all electrodes of subjects A-D, respectively, showed a significant correlation with at least one of the conditions of the articulator movement task. Subject A & B performed this task twice and for the second run 71.88% (46/64) and 92.97% (119/128) of the electrodes were significant for subjects A & B, respectively. For the tongue movement task, 53.12% (34/64), 68.75% (88/128), and 64.84% (83/128) of the electrodes of subjects A, B and D showed a significant correlation with the task. Note that subject C did not perform this task. Subject A & B also performed a run with the tongue outside the mouth and respectively 51.56% (33/64) & 68.75% (88/128) of the electrodes showed a significant correlation with the task. Subject B performed a second run with the tongue inside the mouth and 64.06% (82/128) of the electrodes were significant.

Spatial Classification Accuracy

When classification was based on only spatial features, average classification accuracy for the articulator movements was 92.12% (SD=6.61, n=4) for run 1, and 85.62% (SD=2.65, n=2) for run 2 (Figure 1). For the tongue movements inside the mouth, the classification accuracy was 83.75% (SD=5.73, n=3) for run 1. Only one subject did a second run of this task and had a classification accuracy of 71.25%. For the tongue movements outside the mouth, the mean accuracy was 81.88% (SD=22.10, n=2). Classification was significant ($p < 0.001$) for all subjects, runs and tasks. See supplementary figure S2 for the confusion matrices.

Spatio-Temporal Classification Accuracy

When classification was based on spatio-temporal features, mean classification accuracy was 89.24% (SD=5.62, n=4) for run 1, and 85.62% (SD=2.65, n=2) for run 2 (Figure 1). For the tongue movements inside the mouth the classification accuracy was 74.58% (SD=13.94, n=3) for run 1. The one subject who did a second run had a classification accuracy of 68.75%. For the tongue movements outside the mouth the mean accuracy was 73.75% (SD=15.91, n=2). Classification was significant ($p < 0.001$) for all subjects, runs and tasks. See supplementary figure S3 for the confusion matrices.

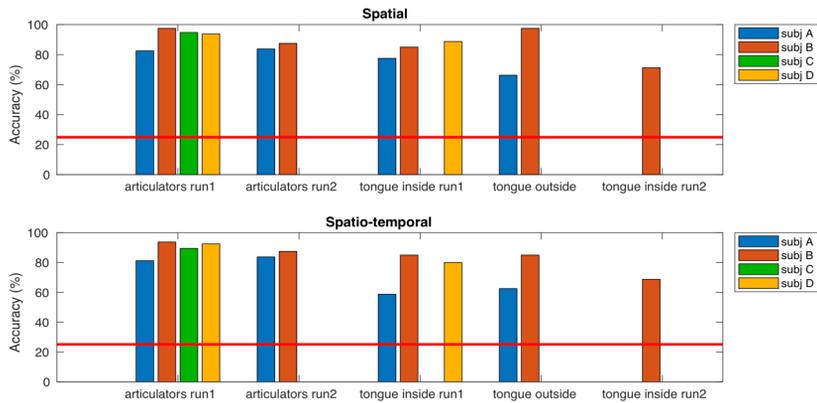


Figure 1. Classification accuracies. The classification accuracies are shown for each subject (A-D, in color), per run separately. Classification was based on all electrodes that showed a significant relation to the task. Runs are indicated on the x-axis and the classification accuracy is indicated on the y-axis. In the upper panel, the spatial classification accuracies are shown, in the lower panel, the accuracies for spatio-temporal classification. The red line indicates the chance level of 25%. Note that not all subjects performed two runs and subject C only performed the articulator movement task.

Anatomical Localization

Since most participants performed only one run of each task, the next results are based on the first run of each participant. For the participants that did perform multiple runs, however, the results of the second run are similar to that of the first, as displayed in the supplementary material (Figures S4 & S5).

The most informative electrodes for classification of articulator movements were more spread out over the sensorimotor cortex compared to the tongue task which showed to have the most informative electrodes more clustered in the central aspects of the vSMC (Figure 2).

The searchlight procedure shows for each cortical area sampling size, the maximum classification accuracy score that can be obtained. For articulator movements, the cortical area necessary for a subject to reach more than 70% accuracy was on average 0.62 square cm (SD=0.40, n=4), and to reach more than 80% was on average 1.18 square cm (SD=0.83, n=4). For tongue movement directions, 0.87 square cm (SD=0.31, n=3) was needed to reach 70% accuracy and 1.05 square cm (SD=0.21, n=2) to reach 80% accuracy, see Figure 3, (note that one subject (A) did not reach 80% for tongue movement directions). These results suggest that the cortical foci (electrodes) that contain information about movements of different articulators or different tongue directions are relatively close together. Figure 4 shows for each electrode the average accuracy of all classifications in which that electrode was involved during the searchlight procedure. This indicated that the cortical area that is involved in the generation of movements is larger for different articulator movements than for different tongue direction movements.

For each articulator movement and tongue movement direction, we investigated where on the sensorimotor cortex the activity was located, see Figure 5 (for individual subject maps see supplementary figure S6). Both for the different articulator movements and the different tongue movement directions, we found much overlap between movements. For articulator movements, there seemed to be a topographical order with respectively the jaw, larynx, tongue, lips and again larynx, oriented in a ventral to dorsal direction. For the different tongue movements, the topographical ordering was less clear (especially in S1) although within M1 the leftward movements seemed to be located somewhat posterior-ventral whereas rightward movements were located more anterior-dorsal. The upward movements seemed to be somewhat more ventral and the downward movements somewhat more dorsal. Note however, that these results are based only on the first run, since only a limited number of participants performed additional runs, and that not all sampled areas were covered by all subjects.

Discussion

In this study, we confirmed that with a very straightforward and robust classification procedure, movements from different articulators can be distinguished from each other using ECoG (Figure 1). Importantly, we found that different movement directions of the tongue can also be classified. Classification of articulator movements and tongue movement directions can be accomplished using only spatial information (spatial classification) and adding time as a feature of the classification (spatio-temporal classification) did not lead to much higher classification accuracies. However, when the start of movements within the same class varies, the mean time trace in the spatio-temporal classification may not lead to more accurate scores than taking the mean over time (spatial classification). Correcting for variations of neural activity onset may potentially lead to even better classification than we currently obtained (see for instance Branco et al. 2017).

Overall, the repeated runs yielded similar results albeit somewhat lower, possibly due to fatigue. Since we were not able to correct for performance as most movements were invisible from the outside, we cannot rule out that task performance might have caused potential differences in classification accuracy between runs or between subjects. For the articulator movements, subject A indicated to sometimes have difficulty making the movements in time, which may explain partly why the scores were somewhat lower for that subject than for the other subjects. However, for the two subjects (A & B) who performed one run of tongue movements outside the mouth, there was no noticeable difference in performance and both participants were very accurate, even though their classification accuracies differ quite a bit. Since we sampled from a smaller cortical area in subject A, it could be that differences in electrode coverage may have caused these classification accuracy differences.

Areas (electrodes) that were most informative for distinguishing different articulators, extended a larger area of the sensorimotor cortex and were more separated from each other than for the tongue task which showed to have the most informative electrodes more clustered in the central aspects of the SMC (Figure 2). Even so, using ECoG grids with closely spaced electrodes, classification could be obtained from areas of approximately 1 square centimetre (Figure 3). This indicates that the areas that contain information about different movements or movement directions

are close together. The area that *can* be used for grid placement leading to accurate classification was more concise for tongue movements than for articulator movements (Figure 4).

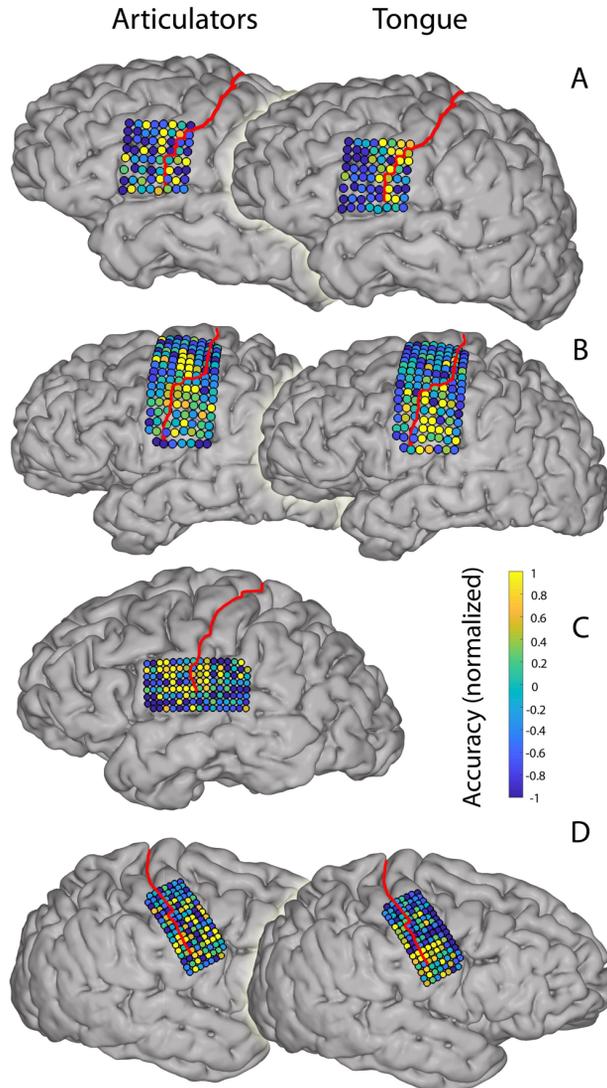


Figure 2. Most informative electrodes. The most informative electrodes are shown in color for each subject (A-D) and for each task. Informative electrodes were determined by a random search procedure. Colors indicate the normalized average accuracy score of this procedure. The warmer the color the higher the classification was on average if that electrode was included. Results are based on the first run. See supplementary material for the data of the other runs.

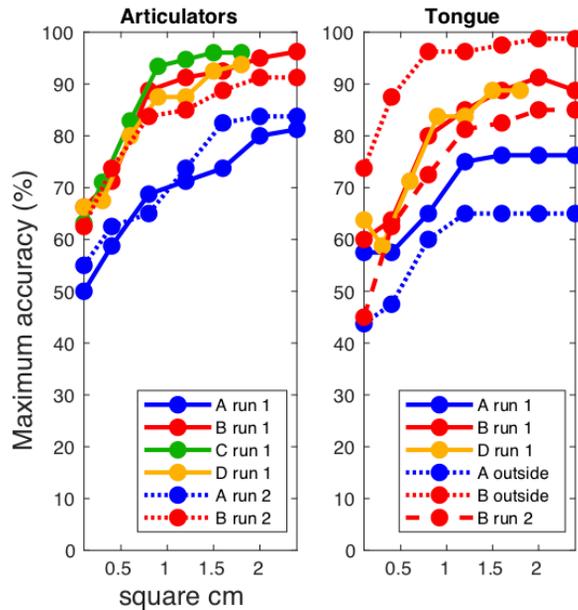


Figure 3. Size of cortical area required for accurate classification. *The maximum classification accuracy (y-axis) is shown for different sizes of cortical surface areas (x-axis), for each subject. This was calculated by using a search light approach in which we increased the sampled area of the search light and indicated where the highest accuracy was.*

Finally, visual inspection of the activity patterns for each articulator movement and tongue movement direction, suggests a topographical ordering, most clearly for the movements of different articulators, with a ventral to dorsal ordering of the jaw, larynx, the tongue, lips and a second area for larynx, respectively (Figure 5).

Relation to Previous Research

The current results are in agreement with a previous fMRI study that showed that lip, tongue, jaw and larynx movements can be distinguished from each other based on sensorimotor cortex activity (Bleichner et al. 2015), albeit with a different method and less cortical coverage (limited to the skull-lining surface). In addition, we showed that different directions of tongue movement can be classified as well from the same region. Classification of different tongue movement directions is in agreement with single cell recordings in primates that have shown different neurons to be related to different movements of the tongue. Stimulation of some neurons, for instance, resulted in tongue protrusion whereas stimulation of other neurons resulted in retraction (Yao et al. 2002). Furthermore, some neurons showed significantly different firing rates for tongue protrusion in different directions (Murray and Sessle 1992). Here we extended those results by showing that different movement directions can also be measured from the SMC in human cortex from surface electrode recordings of neuronal ensembles.

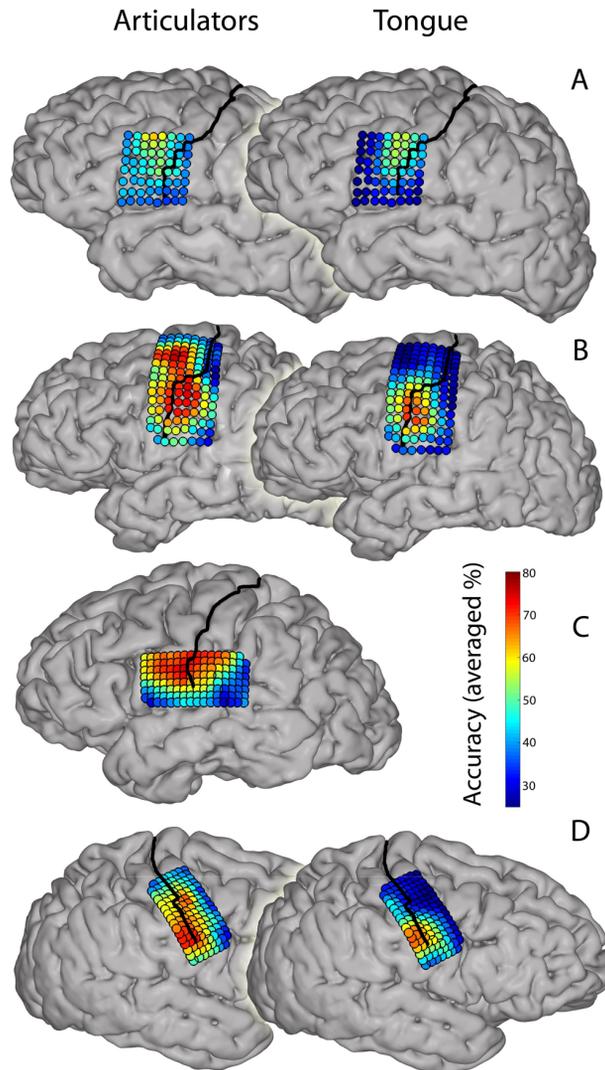


Figure 4. Search light results. The most informative areas are shown in color for each subject (A-D) and for each task. Informative areas were determined by a search light procedure. For this plot, we choose a search light of three electrode rows and columns (a 3x3 grid size), corresponding to approximately 12x12 mm for subject A & B and 9x9 mm for subjects C & D. Colors indicate the average accuracy score of this procedure. Results are based on the first run. See supplementary material for the data of the other runs.

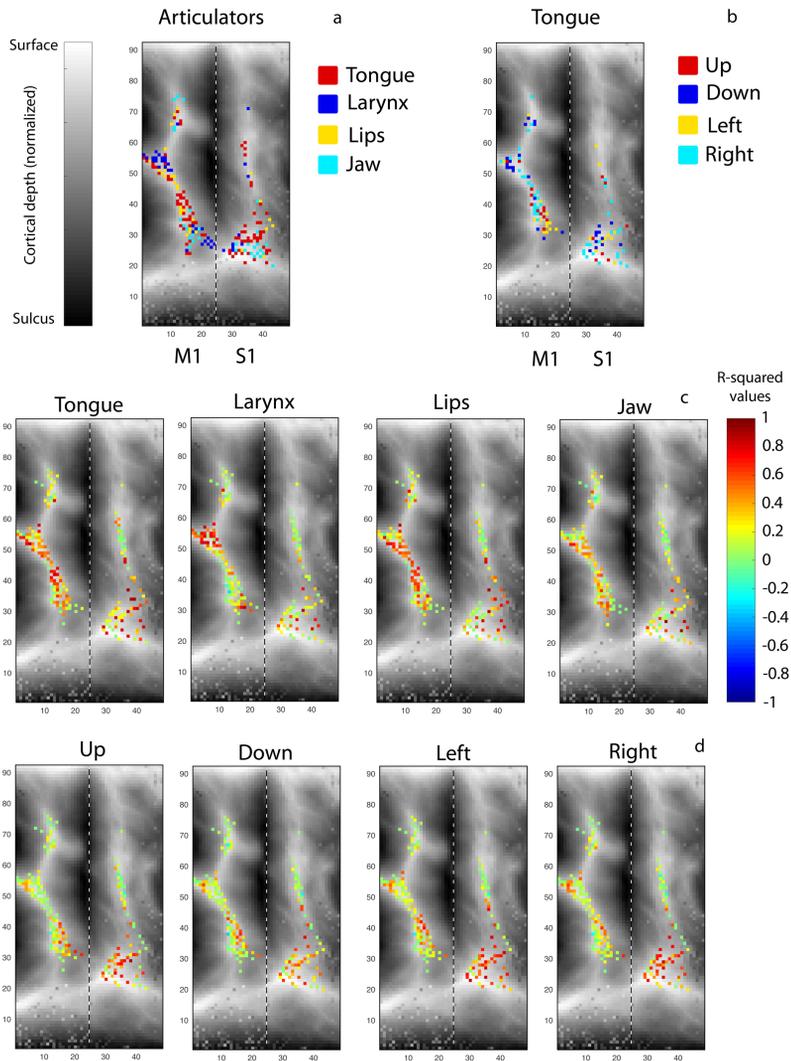


Figure 5. Localization of different movements. In the top panels (a & b), the average localization of different articulator and tongue movements is shown on an inflated and normalized brain surface with darker colors indicating the sulcus and lighter colors indicating the surface (for anatomical reference of these matrices see Bruurmijn et al. 2017). Each colored square indicates one electrode. The left side of each plot indicates the precentral gyrus (M1) and the right side the postcentral gyrus (S1), also for subject D in whom the right hemisphere was recorded. The dashed line indicates the central sulcus. Colors indicate for each electrode which movement had the highest r^2 -value in a winner takes all fashion. For the top panel plots (a & b), electrodes with an r^2 -value less than 0.2 are not shown. The coordinates on the x-axis and y-axis indicate the number of ‘tiles’ in the anterior-posterior and ventral-dorsal direction, respectively. Each tile represents about 1 mm of inflated cortex. In the lower panels (c & d), the average of the r^2 -values per movement are shown. Results are based on the first run. See supplementary Figure S6 for the data per participant for all runs.

Similar as with fMRI and stimulation studies (Penfield and Boldrey 1937; Hesselmann et al. 2004; Grabski et al. 2012) we also found a large overlap in the location of neural activation for the movements of different articulators. However, we found by visual inspection of the activity patterns, a ventral-dorsal topographical ordering in the neural activity hotspots for the different articulator movements, with the jaw and larynx mostly ventral, the tongue superior to that and the lips and a second area for larynx mostly dorsal. Note however, that these results are based on a group map in which not all areas are equally covered by all subjects. These results are similar, however, to those of previous studies with respect to the notion of the two larynx areas (Penfield and Boldrey 1937; Bouchard et al. 2013) and the localization of the tongue in the middle of the vSMC (Penfield and Boldrey 1937; Grabski et al. 2012; Bouchard et al. 2013).

Jaw-related activity, however, was found along the central sulcus with the highest activity mostly ventral, while some other studies found this to be more dorsal to the tongue area (Grabski et al. 2012; Bouchard et al. 2013). This discrepancy may be caused by methodological differences: in our study, people had to clench the jaw, while in the other studies, the jaw was opened instead of closed. Stimulation results (Penfield and Boldrey 1937), indeed, support the notion of jaw activity along the central sulcus, also ventrally, similar to what we found. For the lips, we found a similar location with respect to the ventral-dorsal ordering as previous studies, but the activity hotspot was more in the precentral gyrus in our study, whereas Bouchard et al. (2013) found it to be more in the postcentral gyrus. Again, task differences may have caused this discrepancy: we used lip protrusion whereas Bouchard et al. (2013) used speech movements. It has been suggested that reorganisation of motor activity comes with movement training (Karni et al. 1998; Schieber 2001) and since it can be speculated that lip protrusion is a somewhat less common movement than speech movements, this may lead to more motor related neural activity in the precentral gyrus than speech movements (Saarinen et al. 2006).

For the different tongue movements, the topographical representation was less clear (especially in S1). Similar as for the articulator movements, we did find a large overlap in the neural patterns for different tongue movements. However, although some results may have been driven more by some subjects than by others due to differences in the sampled area (see also supplementary figure S6), there seemed to be a bit of spatial ordering within M1. Leftward movements were located mostly in the posterior ventral areas whereas rightward movements were more located in anterior dorsal areas. Visual inspection suggests that the upward movements were located more ventral and the downward movements somewhat more dorsal. These results seem to parallel those found for hand movements with fMRI (Toxopeus et al. 2011) although such comparisons remain rather subjective. Interestingly, Toxopeus and colleagues (2011) mainly found a distinction of different directions along the depth of the sulcus for hand movements. We found that different directions can also be well distinguished from recordings from the surface of the cortex.

Implications

Interest into the exact functioning of the sensorimotor cortex with respect to speech and articulator movements has grown recently since this area has shown to be a potential target for speech brain computer interface (BCI) systems (e.g. Kellis et al. 2010; Brumberg et al. 2011; Pei et al. 2011;

Mugler et al. 2014; Herff et al. 2015; Ramsey et al. 2018). These are systems that aim to provide severely paralyzed subjects (e.g. locked-in syndrome patients) with an aid for communication (Wolpaw et al. 2002). BCI systems read out specific brain signals that patients have learned to control and which a computer subsequently can recognize and turn into a desired communication-action (e.g. spelling a letter, calling a care giver). Using SMC activity related to (attempted) speech movements is an intuitive and attractive way for decoding speech, since the production of different sounds involves different articulator movements which are thought to lead to different SMC activity patterns that can be recognized and classified by a computer. Here, we show that indeed, a lot of information about articulator movements and tongue movement directions can be extracted from the surface of the SMC, which indeed indicates different neural patterns for different movements and directions.

Previous studies on neural activity based classification of speech units (phonemes, words, sentences) have shown potential for using SMC activity. Isolated phonemes have been classified based on SMC activity for instance during overt speech (4 sounds: $71.9 \pm 8.8\%$ accuracy; Ramsey et al. 2018) and even whole sentences could be reconstructed based on this in natural speech (word accuracy around 75% in a set of 10 words; Herff et al. 2015). However, classification of speech units in larger sets and during attempted speech remains challenging (see e.g. Kellis et al. 2010; Brumberg et al. 2011; Mugler et al. 2014) and more research to the exact functioning of the SMC with respect to articulator movements is therefore needed.

Here we found that information about articulator movements and tongue movement directions can be extracted from both the left and right hemisphere, suggesting that the representation of articulator movements is bilateral, which is in agreement to earlier fMRI studies (Bunton 2008). Furthermore, a fundamental question that is important in the development of BCIs is, which areas need to be covered to record from and what spatial recording resolution and size is needed for these systems to perform accurately. We show that with small electrode grids (ca. 1 square cm), we are able to get high classification results. These results are important for fully implantable systems since only a limited number of electrodes can be recorded with such systems. Furthermore, keeping the invasiveness of such systems as low as possible is also important to prevent possible complications related to surgery (Wong et al. 2009). Obviously, precise grid placement is very essential here. Possibly preoperative fMRI may point to the optimal target location. This method has been successfully used previously for hand movements (Vansteensel et al. 2016) and visual responses (Zhang et al. 2013). Unfortunately, we were not able to test also this task with fMRI in these patients for articulator movements and tongue movement directions.

Other BCI systems that are not directed to speech have also been designed. A system that contrasts for instance SMC activity during attempted hand movement against rest has been shown to be very reliable (Vansteensel et al. 2016). However, this system makes use of simple control signals (1 degree of freedom) which limits the speed of communication (2 letters per minute). Here we showed that a system such as described by Vansteensel and colleagues may be extended with more degrees of freedom by using vSMC activity related to non-speech related articulator and tongue movements. This may potentially increase the control speed of such BCIs. Finally, it is important to note that replication of these experiments with attempted movements is needed since current results might be influenced by, for instance, sensory feedback which is not present for

attempted movements. In addition, it is important to test how the current results relate to real time applications in which subjects get feedback on the classification performance. It has been suggested for instance that neural feedback may improve classification accuracies (Wolpaw et al. 2002; Taylor et al. 2002).

Limitations

In the current study, only a small number of participants were included. More data would have strengthened the current results although even with a small number of participants we were able to show the distinctiveness in SMC neural activity for different articulator movements and different articulator movement directions.

Conclusion

We demonstrate here that with high-density subdural ECoG recordings from the sensorimotor cortex surface, movements of different articulators can be distinguished from each other based on neural activity. In addition, we showed that different movements of one articulator (the tongue) can be classified. Both between and within articulator movements can be classified from a very small cortical area (ca. 1 cm²). These neural signals of articulator and tongue movements may be a new additional control source for BCI applications.

Acknowledgements

This work was supported by the European Union (ERC-Advanced ‘iConnect’ grant 320708). The authors thank the participants for their contribution. Also, they would like to thank the staff of the clinical neurophysiology department and the neurosurgeons.

References

- Bleichner MG, Jansma JM, Salari E, et al (2015) Classification of mouth movements using 7 T fMRI. *J Neural Eng* 12:066026. doi: 10.1088/1741-2560/12/6/066026
- Booij G (1999) *The Phonology of Dutch*. Clarendon Press
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. *Nature* 495:327–332. doi: 10.1038/nature11911
- Branco MP, Freudenburg ZV, Aarnoutse EJ, et al (2017) Decoding hand gestures from primary somatosensory cortex using high-density ECoG. *NeuroImage* 147:130–142. doi: 10.1016/j.neuroimage.2016.12.004
- Branco MP, Freudenburg ZV, Aarnoutse EJ, et al (2018a) Optimization of sampling rate and smoothing improves classification of high frequency power in electrocorticographic brain signals. *Biomed Phys Eng Express* 4:045012. doi: 10.1088/2057-1976/aac3ac
- Branco MP, Gaglianese A, Glen DR, et al (2018b) ALICE: A tool for automatic localization of intra-cranial electrodes for clinical and high-density grids. *J Neurosci Methods* 301:43–51. doi: 10.1016/j.jneumeth.2017.10.022
- Branco MP, Leibbrand M, Vansteensel MJ, et al (2018c) GridLoc: An automatic and unsupervised localization method for high-density ECoG grids. *NeuroImage* 179:225–234. doi: 10.1016/j.neuroimage.2018.06.050
- Brumberg JS, Wright EJ, Andreasen DS, et al (2011) Classification of Intended Phoneme Production from Chronic Intracortical Microelectrode Recordings in Speech-Motor Cortex. *Front Neurosci* 5:65. doi: 10.3389/fnins.2011.00065
- Bruns A (2004) Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? *J Neurosci Methods* 137:321–332. doi: 10.1016/j.jneumeth.2004.03.002
- Bruurmijn MLCM, Pereboom IPL, Vansteensel MJ, et al (2017) Preservation of hand movement representation in the sensorimotor areas of amputees. *Brain* 140:3166–3178. doi: 10.1093/brain/awx274
- Bunton K (2008) Speech versus Nonspeech: Different Tasks, Different Neural Organization. *Semin Speech Lang* 29:267–275. doi: 10.1055/s-0028-1103390
- Chartier J, Anumanchipalli GK, Johnson K, Chang EF (2018) Encoding of Articulatory Kinematic Trajectories in Human Speech Sensorimotor Cortex. *Neuron*. doi: 10.1016/j.neuron.2018.04.031
- Crone NE, Miglioretti DL, Gordon B, Lesser RP (1998) Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain* 121:2301–2315. doi: 10.1093/brain/121.12.2301
- Georgopoulos, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J Neurosci* 2:1527–1537
- Grabski K, Lamalle L, Vilain C, et al (2012) Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Hum Brain Mapp* 33:2306–2321. doi: 10.1002/hbm.21363
- Guenther FH, Hickok G (2015) Chapter 9 - Role of the auditory system in speech production. In: Aminoff MJ, Boller F, Swaab DF (eds) *Handbook of Clinical Neurology*. Elsevier, pp 161–175
- Herff C, Heger D, de Pestiers A, et al (2015) Brain-to-text: decoding spoken phrases from phone representations in the brain. *Neural Technol* 9:217. doi: 10.3389/fnins.2015.00217
- Hermes D, Miller KJ, Noordmans HJ, et al (2010) Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J Neurosci Methods* 185:293–298. doi: 10.1016/j.jneumeth.2009.10.005
- Hesselmann V, Sorger B, Lasek K, et al (2004) Discriminating the Cortical Representation Sites of Tongue and Lip Movement by Functional MRI. *Brain Topogr* 16:159–167. doi: 10.1023/B:BRAT.0000019184.63249.e8
- Karni A, Meyer G, Rey-Hipolito C, et al (1998) The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proc Natl Acad Sci* 95:861–868. doi: 10.1073/pnas.95.3.861
- Kellis S, Miller K, Thomson K, et al (2010) Decoding spoken words using local field potentials recorded from the cortical surface. *J Neural Eng* 7:056007. doi: 10.1088/1741-2560/7/5/056007
- Levelt WJM (1993) *Speaking: From Intention to Articulation*. MIT Press
- Manning JR, Jacobs J, Fried I, Kahana MJ (2009) Broadband shifts in LFP power spectra are correlated with single-neuron spiking in humans. *J Neurosci Off J Soc Neurosci* 29:13613. doi: 10.1523/JNEUROSCI.2041-09.2009
- Meister IG, Wilson SM, Deblieck C, et al (2007) The Essential Role of Premotor Cortex in Speech Perception. *Curr Biol* 17:1692–1696. doi: 10.1016/j.cub.2007.08.064
- Mermelstein P (1973) Articulatory model for the study of speech production. *J Acoust Soc Am* 53:1070–1082. doi: 10.1121/1.1913427
- Miller KJ, Sorensen LB, Ojemann JG, Nijss M den (2009) Power-Law Scaling in the Brain Surface Electric

- Potential. *PLOS Comput Biol* 5:e1000609. doi: 10.1371/journal.pcbi.1000609
- Moran DW, Schwartz AB (1999) Motor Cortical Representation of Speed and Direction During Reaching. *J Neurophysiol* 82:2676–2692
- Mugler EM, Patton JL, Flint RD, et al (2014) Direct classification of all American English phonemes using signals from functional speech motor cortex. *J Neural Eng* 11:035015. doi: 10.1088/1741-2560/11/3/035015
- Murray GM, Sessle BJ (1992) Functional properties of single neurons in the face primary motor cortex of the primate. III. Relations with different directions of trained tongue protrusion. *J Neurophysiol* 67:775–785. doi: 10.1152/jn.1992.67.3.775
- Pei X, Barbour DL, Leuthardt EC, Schalk G (2011) Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. *J Neural Eng* 8:046028. doi: 10.1088/1741-2560/8/4/046028
- Penfield W, Boldrey E (1937) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain J Neurol* 60:389–443. doi: 10.1093/brain/60.4.389
- Ramsey NF, Salari E, Aarnoutse EJ, et al (2018) Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage*. doi: 10.1016/j.neuroimage.2017.10.011
- Ray S, Maunsell JHR (2011) Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLOS Biol* 9:e1000610. doi: 10.1371/journal.pbio.1000610
- Rietveld A, van Heuven V (2001) *Algemene Fonetiek*, 2nd edn. Coutinho, Bussum
- Saarinén T, Laaksonen H, Parviainen T, Salmelin R (2006) Motor Cortex Dynamics in Visuomotor Production of Speech and Non-speech Mouth Movements. *Cereb Cortex* 16:212–222. doi: 10.1093/cercor/bhi099
- Salari E, Freudenburg ZV, Vansteensel MJ, Ramsey NF (2018a) Spatial-Temporal Dynamics of the Sensorimotor Cortex: Sustained and Transient Activity. *IEEE Trans Neural Syst Rehabil Eng* 26:1084–1092. doi: 10.1109/TNSRE.2018.2821058
- Salari E, Freudenburg ZV, Vansteensel MJ, Ramsey NF (2018b) The influence of prior pronunciations on sensorimotor cortex activity patterns during vowel production. *J Neural Eng* 15:066025. doi: 10.1088/1741-2552/aae329
- Schieber MH (2001) Constraints on Somatotopic Organization in the Primary Motor Cortex. *J Neurophysiol* 86:2125–2143
- Taylor DM, Tillery SIH, Schwartz AB (2002) Direct Cortical Control of 3D Neuroprosthetic Devices. *Science* 296:1829–1832. doi: 10.1126/science.1070291
- Toxopeus CM, Jong BM de, Valsan G, et al (2011) Direction of Movement Is Encoded in the Human Primary Motor Cortex. *PLOS ONE* 6:e27838. doi: 10.1371/journal.pone.0027838
- Vansteensel MJ, Pels EGM, Bleichner MG, et al (2016) Fully Implanted Brain–Computer Interface in a Locked-In Patient with ALS. *N Engl J Med* 375:2060–6. doi: 10.1056/NEJMoa1608085
- Wang J, Green JR, Samal A (2013) Individual articulator’s contribution to phoneme production. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp 7785–7789
- Wolpaw JR, Birbaumer N, McFarland DJ, et al (2002) Brain–computer interfaces for communication and control. *Clin Neurophysiol* 113:767–791. doi: 10.1016/S1388-2457(02)00057-3
- Wong CH, Birkett J, Byth K, et al (2009) Risk factors for complications during intracranial electrode recording in presurgical evaluation of drug resistant partial epilepsy. *Acta Neurochir (Wien)* 151:37. doi: 10.1007/s00701-008-0171-7
- Yao D, Yamamura K, Narita N, et al (2002) Neuronal Activity Patterns in Primate Primary Motor Cortex Related to Trained or Semiautomatic Jaw and Tongue Movements. *J Neurophysiol* 87:2531–2541. doi: 10.1152/jn.2002.87.5.2531
- Zhang D, Song H, Xu R, et al (2013) Toward a minimally invasive brain–computer interface using a single subdural channel: A visual speller study. *NeuroImage* 71:30–41. doi: 10.1016/j.neuroimage.2012.12.069

Appendix

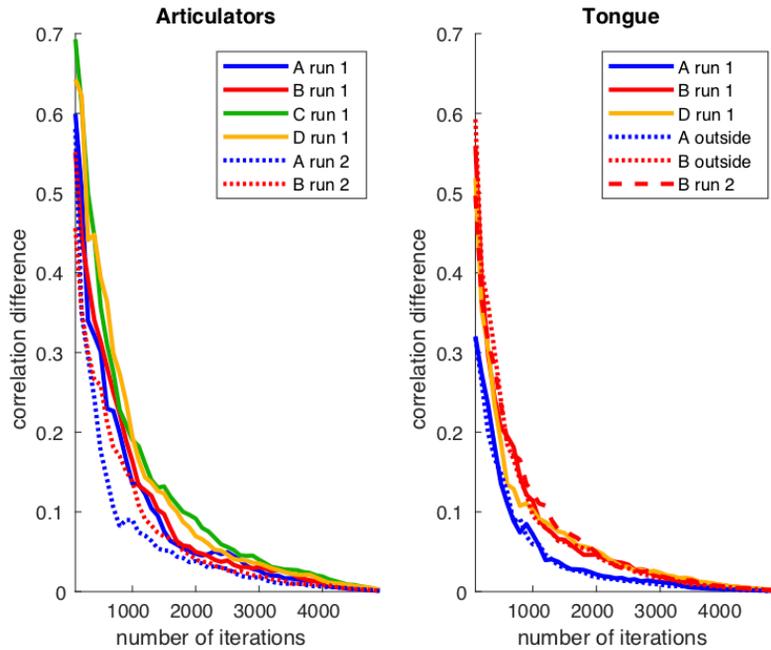


Figure S1. Random search stabilization. *The stabilization of the random search result in relation to the number of iterations (x-axis) is shown for each participant and run. For each iteration, the pattern of most informative electrodes is calculated. On the y-axis, the average difference in correlation of each pattern with the subsequent patterns is shown.*

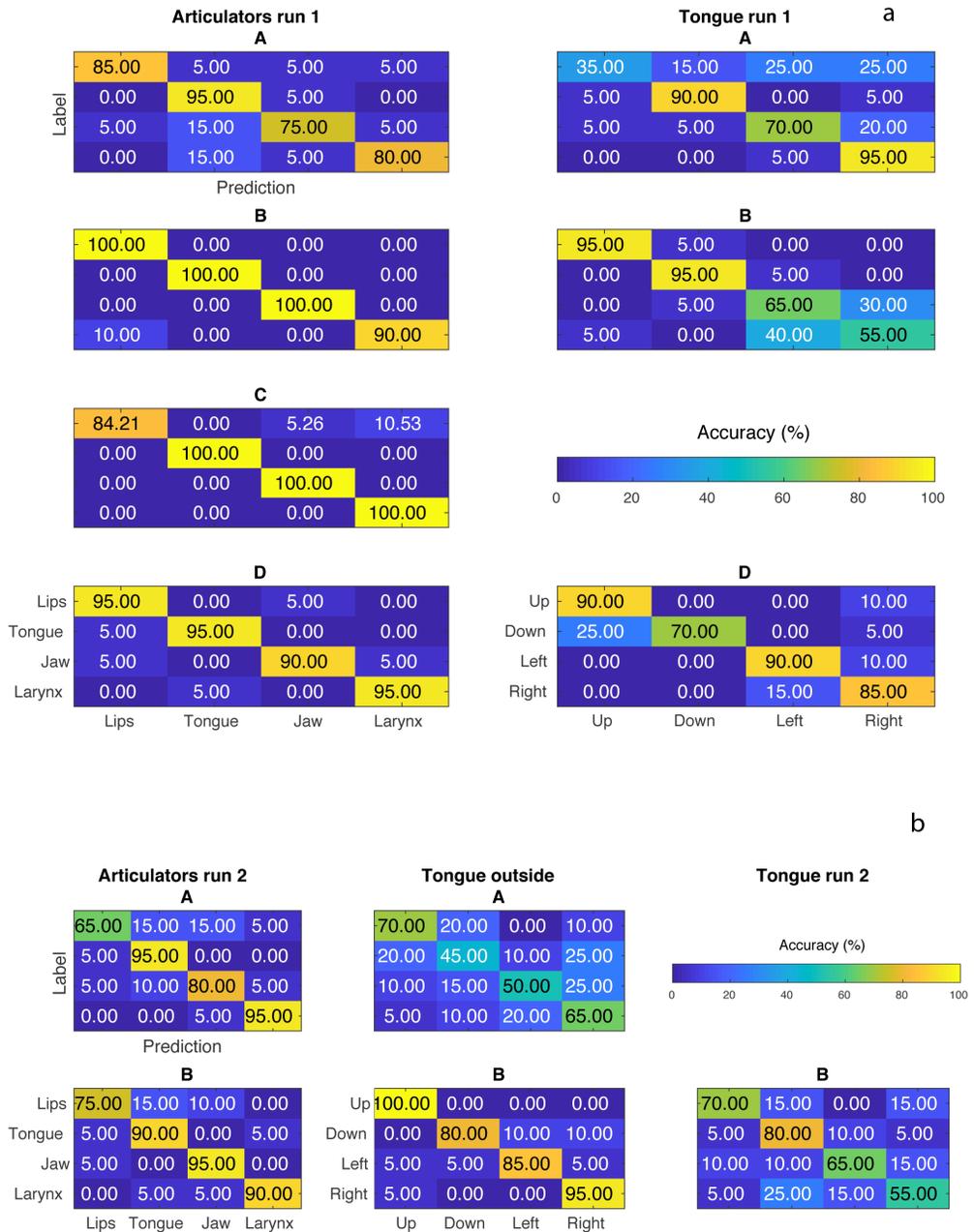


Figure S2. Confusion matrices of the spatial classification. The confusion matrixes are shown for each subject of the first run (a) and for the subjects that performed additional runs (b). Colors indicate the percentage of trials classified within each class.

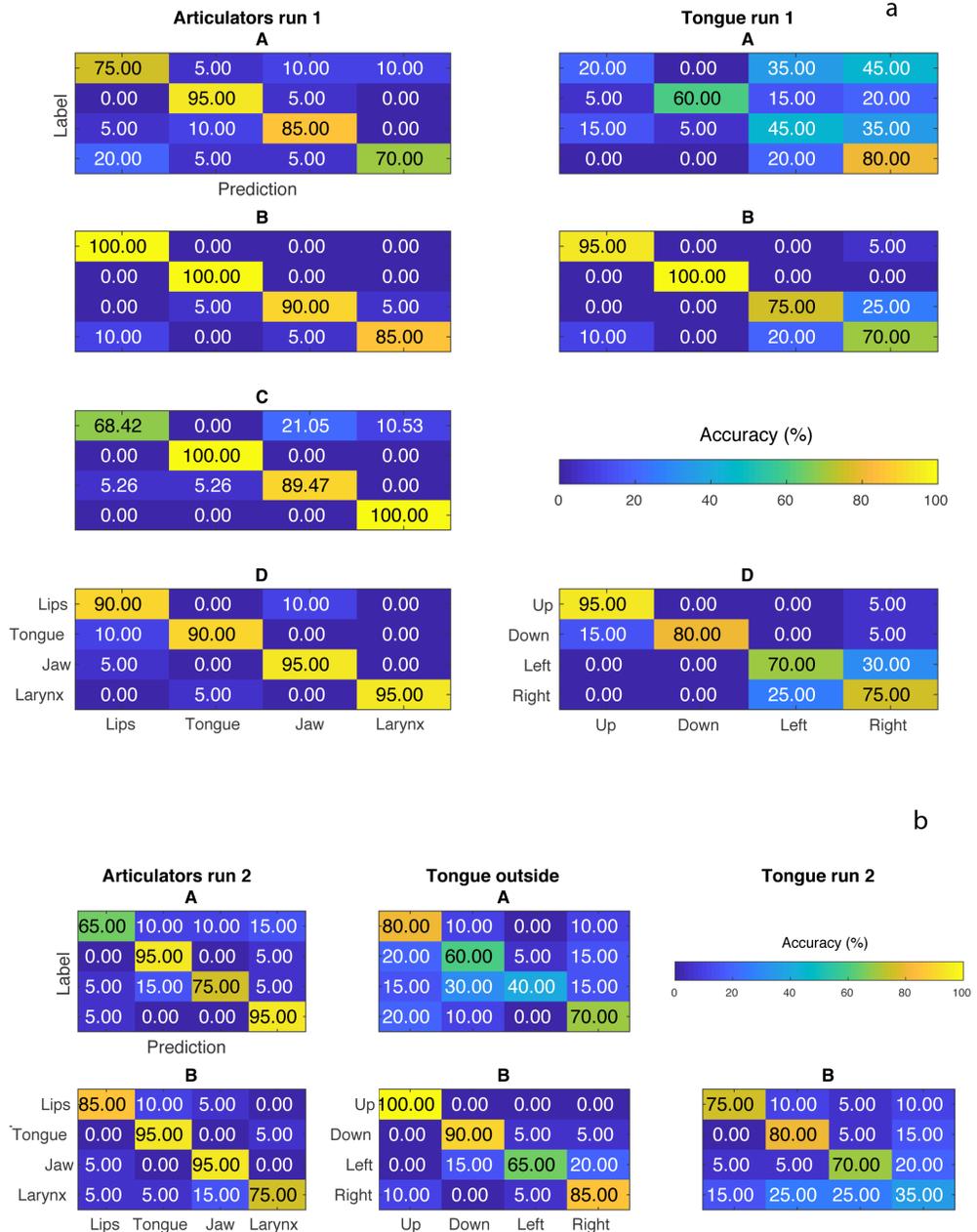


Figure S3. Confusion matrices of the spatio-temporal classification. The confusion matrixes are shown for each subject of the first run (a) and for the subjects that performed additional runs (b). Colors indicate the percentage of trials classified within each class.

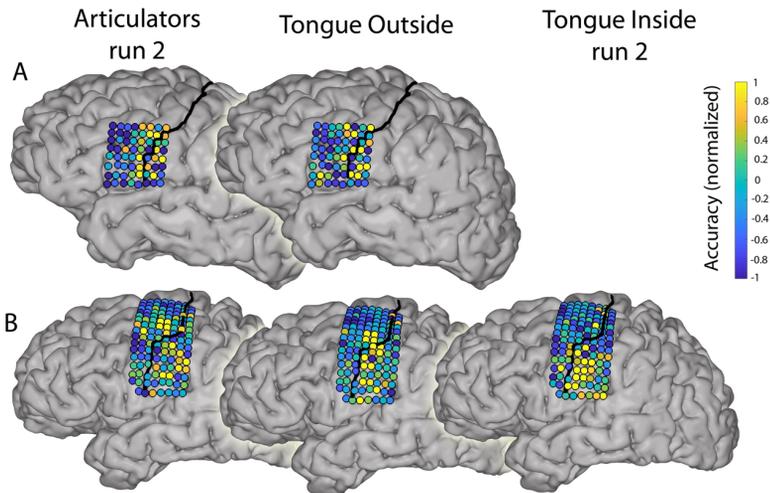


Figure S4. Most informative electrodes for additional runs. *The most informative electrodes are shown in color for each subject (A& B) and for each task. Informative electrodes were determined by a random search procedure. Colors indicate the normalized average accuracy score of this procedure. The warmer the color the higher the classification was on average if that electrode was included.*

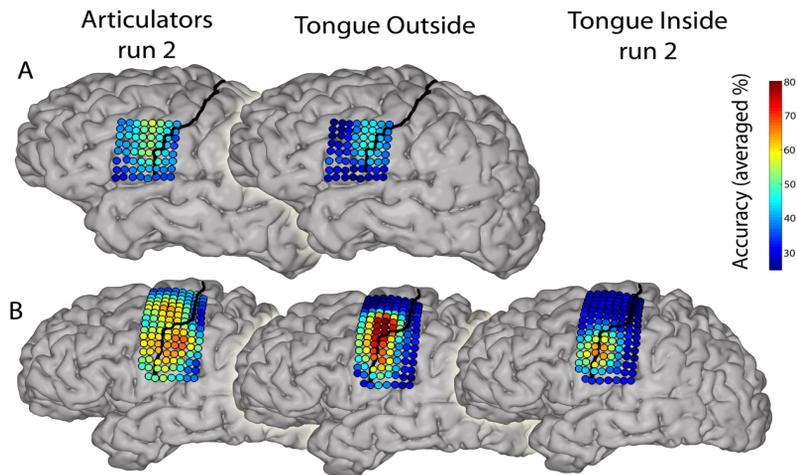


Figure S5. Search light results for additional runs. *The most informative areas are shown in color for each subject (A& B) and for each task. Informative areas were determined by a search light procedure. For this plot, we choose a search light of three electrode rows and columns (a 3x3 grid size), corresponding to approximately 12x12 mm for subject A & B. Colors indicate the average accuracy score of this procedure.*

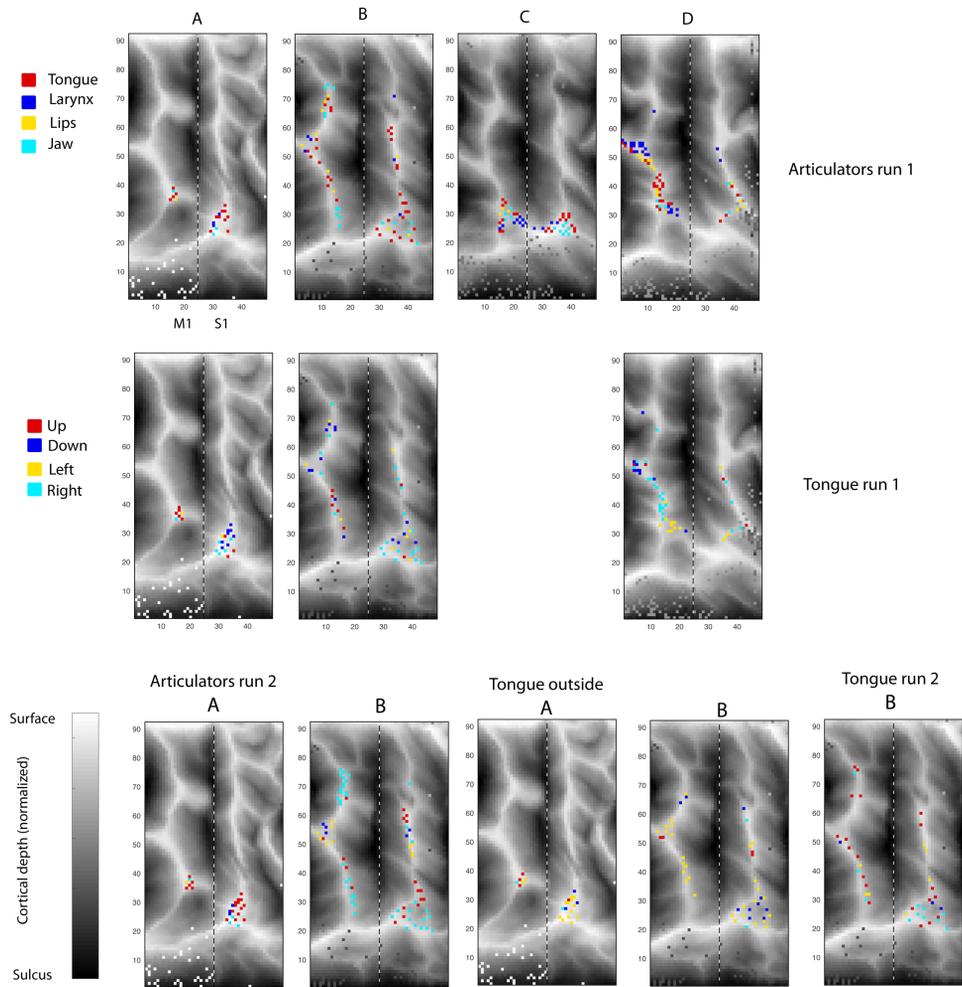


Figure S6. Localization of different movements per subject and per run. Each colored square indicates one electrode on an inflated and normalized brain surface. Darker colors indicating the sulcus and lighter colors indicating the surface (for anatomical reference of these matrices see Bruurmijn et al. (2017)). The left side of each plot indicates the precentral gyrus (M1) and the right side the postcentral gyrus (S1). The dashed line indicates the central sulcus. Colors indicate for each electrode which movement had the highest r^2 -value in a winner takes all fashion. Electrodes with an r^2 -value less than 0.2 are not shown. The coordinates on the x-axis and y-axis indicate the number of 'tiles' in the anterior-posterior and ventral-dorsal direction, respectively. Each tile represents about 1 mm of inflated cortex.



Chapter 5

*“Beneath the noise, below the din, I hear a voice, it’s whispering.
In science and in medicine, I was a stranger, you took me in.”*

Chapter 5

Classification of Facial Expressions for Emotional Communication Using BCI

E. Salari, Z.V. Freudenburg, M.J. Vansteensel, N.F. Ramsey

Submitted Article

Abstract

Background: Facial expressions are important for fast communication of one's emotional state without the involvement of speech. For people suffering from severe paralysis, however, the ability of fast communication of emotions is seriously impaired. Brain-computer interfaces have been proposed to assist these people with linguistic communication but are generally cumbersome and restrictive when it comes to the ability to express emotions. **Methods:** In the current study, we investigated the feasibility of providing a fast and intuitive channel for a BCI to convey emotional content by decoding facial expressions. We recorded subdural electrocorticographic (ECoG) activity over the sensorimotor areas in two epilepsy patients who were implanted with ECoG grids and who performed four different facial expressions (happy, sad, surprised and disgusted). We classified which expression participants made, based on the sensorimotor cortex (SMC) activity patterns. **Results:** The accuracy scores were on average 78% across two participants and the most informative areas were located in the aspect of the SMC just below the hand areas. **Discussion:** These findings indicate that emotional expressions can be identified from sensorimotor cortex to a high degree, from a region that is also targeted for implantable speech BCI. This may provide a natural channel for the non-verbal communication of emotions through BCI.

Keywords; Sensorimotor Cortex, Facial Expressions, BCI, ECoG, Emotion

Introduction

Locked-in syndrome (LIS) is a condition where people suffer from (almost) complete paralysis while their mental capacity remains intact (Smith and Delargy, 2005), and where communication through the normal means of vocalization and body language is completely lost. Technology is available to provide people with LIS with a means of communication through residual muscle activations, such as eye movements, but these systems are not usable when all movement is lost. A promising alternative approach is a brain-computer interface (BCI), which enables people to communicate without the involvement of any muscle activity, by directly recording brain signals (Wolpaw et al., 2002) and translating these signals into a control signal for a computer.

Several BCI studies have focused on re-establishing communication through spelling of letters, words and sentences (Brumberg et al., 2011; Herff et al., 2015; Ramsey et al., 2018) and recently Vansteensel et al. (2016) showed that a fully implantable BCI system can provide a reliable means of communication with a high degree of accuracy, for people with LIS in their home environment. The speed of this type of linguistic communication with BCIs is, however, substantially lower (typically several letters per minute; e.g. Vansteensel et al. 2016) than in normal speech (120-180 words per minute; Miller et al., 1976). This obviously prohibits the ability of people with LIS to engage in natural conversations with others.

Facial expressions provide a means of non-linguistic communication (Ekman, 1989; Ekman and Friesen, 2003), allowing people to robustly and quickly convey their emotions without using letters or words. Although people with LIS can't make deliberate facial expressions, attempts to do so may well be subject to identification by means of brain signal decoding technology that exploits the somatotopic cortical organization, which would constitute restoration of the ability to convey emotions. Indications for feasibility of such an 'emotional' BCI is sparse but present in literature. For one, it has been shown that isolated facial movements (e.g. articulator movements; Bleichner et al., 2015; Salari et al., 2018a) can be distinguished based on neural activity in the sensorimotor cortex (SMC). Facial expressions have also been classified based on simultaneous electroencephalography (EEG) and electromyography (EMG) recordings (Chin et al., 2008). It remains unclear, however, whether facial expressions, which require the integration of different facial movements at the same time (Ekman and Friesen, 2003; Cattaneo and Pavesi, 2014), can be distinguished from one another based solely on neural activity. In the current study, as a first step towards an 'emotional' BCI, we analysed neural activity in sensorimotor cortex related to facial expressions, using subdural electrocorticographic signals in 2 human participants.

Methods

Participants

Two participants (A & B) participated (age 30 and 41 years, female) after giving written informed consent. They underwent surgery for epilepsy treatment in the University Medical Center Utrecht, during which subdural ECoG electrodes were temporarily implanted for clinical purposes. For research purposes only, and with written informed consent of the patients, an additional high-density electrode grid of 128 electrodes was placed over the SMC (which had no clinical relevance).

Electrodes had a diameter of 1.17 mm (participant A) or 1 mm (participant B). Inter-electrode distance was 4- or 3-mm, respectively. Only the high-density electrodes were used for the current analysis.

The Ethics Committee of the University Medical Center Utrecht approved this study, which was carried out in accordance with the Declaration of Helsinki (2013).

Task

We examined four different facial expressions (happy, sad, surprised and disgusted), based on their distinction from each other with respect to the required muscles and because they can be accurately distinguished from one another based on video-analysis (Ekman, 1989; Ekman and Friesen, 2003; Taner Eskil and Benli, 2014). Participants were to mimic the facial expressions as shown by pictures of people (Cohn-Kanade data base; Kanade et al., 2000; Lucey et al., 2010) presented on a computer screen approximately 1 m from the participants. Since neural changes in the SMC could be induced by visual perception of actions (Pineda, 2008), posing a confound in decoding, we also in separate trials used written descriptions to instruct participants about which facial expression had to be made. Only electrodes that showed activity to both the pictures and the words were included in further analysis, thereby excluding electrodes that were selective to mode of instruction.

Trials started with a 2000 ms long visual cue (either a picture or a word). The participants were instructed to quickly make the expression and hold it for as long as the cue was visible. The inter-trial interval was 3 seconds. Each facial expression was randomly repeated 5 times in pictures and 5 times in words. Furthermore, we included 10 rest trials, ('-' symbol), which were interleaved with the facial expression trials and which instructed the participants to keep a neutral face.

Data Acquisition & Preprocessing

ECoG signals were recorded at a sampling frequency of 2000Hz (Blackrock Microsystems LLC, Salt Lake City, USA). Electrodes with flat or noisy signal were removed from further analysis, see Salari et al. (2018b).

High-frequency-band (HFB; 65-130 Hz) power changes, which are known to be linked to neuronal firing (Miller, 2010) in the SMC, were used to classify the four different facial expressions. Data analysis was performed using Matlab software (The Mathworks, Inc., Natick, MA, USA). We used a Gabor wavelet dictionary (Bruns, 2004) to calculate for each electrode the HFB power per sample point, see Salari et al. (2018b). We smoothed this signal with a moving average window of 0.5 seconds (Branco et al., 2018a). Subsequently, the r^2 -value of every included electrode was calculated by contrasting rest and active periods for each facial expression separately. Electrodes that were significantly correlated to one or more of the facial expressions were used for classification. Significance level was computed by a Monte Carlo procedure with 10,000 permutations of shuffling trial and rest labels (alpha=0.05, Bonferroni corrected for the number of electrodes). Electrode positions on the brain were identified with ALICE (Branco et al., 2018b) using a post-operative CT scan and a pre-operative MRI scan.

Classification Procedure

Trials were epoched between cue onset and offset. For each trial and each electrode, the mean HFB activity was calculated by taking the average power amplitude over all sample points within the cue period, yielding a spatial HFB activity pattern per trial. This pattern was used as feature for classification of trials on the 4 facial expressions ('classes'), using leave-one-trial-out cross validation. A template-matching-classification-method was used in which each trials' activity pattern was compared to the mean activity pattern of all other trials within a class by means of correlation, which has been used successfully before (e.g. Ramsey et al., 2018). Electrode selection for this procedure was determined by selecting the significant electrodes in the training set (not including the test trial). The significance level of the classification accuracy was determined by using a Monte Carlo procedure, shuffling the labels for classification a 1000 times.

Informative Area Locations and Sizes

We determined which particular regions were most informative for classification against the background of considering BCI applications and minimizing risks of ECoG grid implants (Wong et al., 2009). For this purpose, we used a random search procedure, the methods of which are described in Salari et al. (2018a). In short, random combinations of electrodes were chosen 5000 times and subsequently used for classification. The average accuracy of each electrode was then determined and z-scored. This procedure indicates which electrodes are the most informative, given that any combination of electrodes is possible. In addition, we used a search light procedure (Salari et al., 2018a) to investigate what the classification accuracy would be if we used a smaller (but restricted) portion of the grid. This was also used to determine for different grid sizes what would be the maximum possible accuracy over all investigated grid locations. In addition, we investigated which grid size yielded the highest accuracy, in order to find the minimally required grid size for maximal accuracy.

Results

Task Performance and Electrode Selection

Participants performed the task well, although for participant B some expressions, e.g. surprise, were noted to be rather subtle. The participant did report, however, that the movements she made represented a surprised face for her and we did therefore not exclude any trials. Of all electrodes, 83% (106/128, participant A) and 80% (103/128, participant B), showed a significant correlation with at least one of the facial expressions (for both picture and word trials).

Spatial Classification Accuracy

The classification accuracy was 85% and 70% for participants A and B respectively ($M=77.50$, $SD=10.61$, $n=2$), see Figure 1. Classification was significant ($p<0.001$) for both participants. Surprise was identified the best for both participants. For participant A, all other expressions were equally well recognized. For participant B, sad was also well recognized although happy and disgust were harder to identify, with the former often mistaken by surprise or disgust and the later

by surprise and sad.

Anatomical Localization and Grid Sizes

The most informative electrodes were typically located in the aspect of the sensorimotor cortex just below the hand area (Figure 2). The minimum required grid size to obtain maximum accuracy was 1.6 cm² and 1.2 cm² for participant A and B respectively (M=1.40; Figure 3).

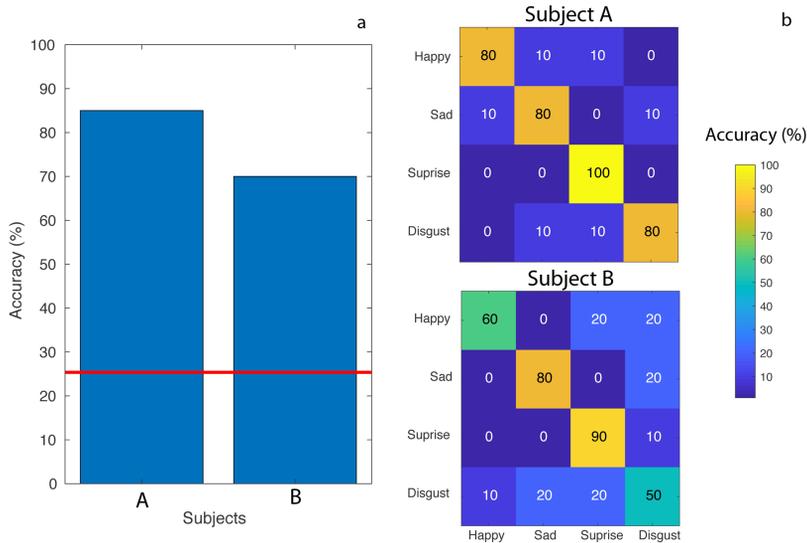


Figure 1. Classification accuracies and confusion matrices. Panel (a) shows the classification accuracies (y-axis) for both participants (x-axis). All electrodes that showed a significant HFB power activation during at least one of the expressions were used for this classification, with the condition that they were active for both stimulus types. The red line represents chance level. In panel (b), the confusion matrices are shown for participant A and B, respectively, with on the y-axis the actual expression, and on the x-axis the predicted expression.

Discussion

We investigated if neural activity in the sensorimotor cortex can be used to distinguish facial expressions. We showed that 4 facial expressions can be classified well, based on SMC neural activity of the left and the right hemisphere (participant A and B, respectively). The dorsolateral aspect of the SMC, an area just below the hand area including both the pre- and postcentral gyrus, was most informative for classification. Furthermore, we found that the grid size had to be on average 1.4 cm² to reach best performance.

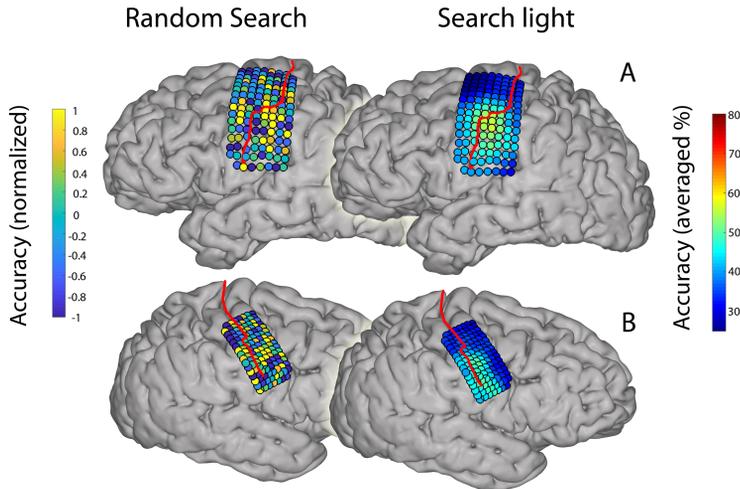


Figure 2. Informative areas. The most informative electrodes are shown on the left for both participants (A-B). The red line indicates the central sulcus. Colors indicate level of attributed information for each electrode, which was determined by a random search procedure. Warm colors indicate high classification accuracies. On the right, the results of the search light procedure are shown, with the most informative areas shown in warm colors. We used a 3x3 electrode configuration for the creation of the current plot (three electrode rows and columns), corresponding to approximately 12x12 mm for participant A and 9x9 mm for participant B. Colors represent, for each electrode, the accuracy score averaged over all configurations that holds that electrode.

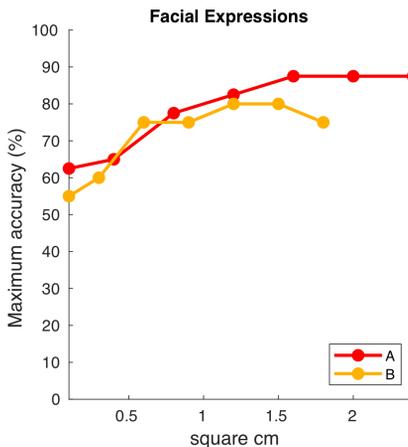


Figure 3. Relation between grid size and classification accuracy. For both participants, the maximum classification accuracy (from all grid locations in the search light) is indicated on the y-axis per size of cortical surface, which is indicated on the x-axis. Results were obtained using the search light approach, in which, for each sampled area of cortex, the classification accuracy was determined. Note that the number of electrodes of each consecutive data-point in this plot is equal for both participants but due to resolution differences the covered surfaces differ.

Importantly, since people had to hold the facial expressions for only two seconds, we show that classification of facial expressions can be established relatively fast, considering the time it takes to spell out letters or words with current home base BCI applications (Vansteensel *et al.*, 2016).

Our results add a new conceptual dimension to previous studies that classified isolated articulator movements (Bleichner *et al.*, 2015; Salari *et al.*, 2018a), by showing that facial expressions of emotion, which entail simultaneous movements of multiple effectors (lips, jaw, nose, eye brows), can be identified to a high degree on neural activity alone. The current results suggest that especially ‘surprised’ was easily recognized which may be explained by the involvement of the jaw for opening of the mouth, which was more extensive for ‘surprised’ compared to the other expressions.

We found that both the left and the right hemisphere could be used for classification of facial expressions, which corresponds to earlier demonstrations of bilateral activation for producing a willful smile (Iwase *et al.*, 2002). Furthermore, the location of most informative electrodes in the current study corresponds to the most active areas found for individual articulator movements (Grabski *et al.*, 2012) and phoneme production (Ramsey *et al.*, 2018). The latter seems to extend further ventrally, however, which may be interesting for future attempts of classifying both linguistic features and facial expressions with the same coverage.

Implications

Our results suggest that a fast way of communicating an emotional state is within reach for BCI systems that employ SMC neural activity. We believe that the ability to quickly express a feeling and add emotional content to BCI output, is of substantial value for people with LIS and may improve participation of LIS patients in conversations. During human social interaction, facial expressions convey someone’s emotional state without the need for language, which is thought to be universal across cultures and may be evolutionary beneficial since they give us an indication of someone’s future behaviour (Ekman, 1989). Positive expressions for instance may enhance cooperation and signal the willingness to interact with others whereas a lack of this may signal the opposite (Schmidt and Cohn, 2001). In addition, facial expressions may add social cues to spoken statements or written text to disambiguate these. Think for instance of text messages (e.g. a joke) that can only be understood through the addition of emoticons. Facial expressions contain therefore extra value to linguistic communication. Furthermore, not only do facial expressions signal information from the speaker to the listener but also vice versa, which enables the speaker to keep track of the listener’s understanding of what is being said. Adding the ability to express a feeling or emotion without speech may therefore improve current BCIs.

Limitations & Future Directions

In the current study we investigated executed movements produced by able-bodied people and it remains unclear if attempted facial expressions made by people with LIS can be accurately classified as well. The most informative locations were found pre- as well as postcentral in both participants. The somatosensory feedback present in actual movement could benefit classification. Yet, attempted movement has been shown to also generate activity in somatosensory cortex

(potentially by feed-forward processing; Bruurmijn *et al.*, 2017) which appears before movement onset and has been shown to provide movement-selective information (Branco *et al.*, 2017). Accordingly, research with upper limb amputees shows a good correspondence between actual and attempted movements (Bruurmijn *et al.*, 2017).

Another limitation is the small number of participants. We do believe, however, that the results are promising and encourage further investigations into facial expressions for BCI. Finally, future research may add more facial expressions and investigate if they can be classified alongside linguistic BCI applications.

Conclusion

We demonstrate that facial expressions can be distinguished from each other based on SMC activity. These results suggest that emotional communication in the form of facial expressions can be added to communication BCIs and potentially increase the usability of neuroprosthetic communication devices.

Acknowledgements

We would like to thank the participants and the clinical staff for their contribution to this experiment. In addition, we would like to thank Philippe Cornelisse, Andreas Wolters and Joesje Kist for their help and advice with respect to the development of the task and their suggestions for relevant literature.

References

- Bleichner MG, Jansma JM, Salari E, Freudenburg ZV, Raemaekers M, Ramsey NF. Classification of mouth movements using 7 T fMRI. *J. Neural Eng.* 2015; 12: 066026.
- Branco MP, Freudenburg ZV, Aarnoutse EJ, Bleichner MG, Vansteensel MJ, Ramsey NF. Decoding hand gestures from primary somatosensory cortex using high-density ECoG. *NeuroImage* 2017; 147: 130–142.
- Branco MP, Freudenburg ZV, Aarnoutse EJ, Vansteensel MJ, Ramsey NF. Optimization of sampling rate and smoothing improves classification of high frequency power in electrocorticographic brain signals. *Biomed. Phys. Eng. Express.* 2018a; 4: 045012.
- Branco MP, Gaglianesi A, Glen DR, Hermes D, Saad ZS, Petridou N, et al. ALICE: A tool for automatic localization of intra-cranial electrodes for clinical and high-density grids. *Journal of Neuroscience Methods* 2018b; 301: 43–51.
- Brumberg JS, Wright EJ, Andreasen DS, Guenther FH, Kennedy PR. Classification of Intended Phoneme Production from Chronic Intracortical Microelectrode Recordings in Speech-Motor Cortex. *Front Neurosci* 2011; 5: 65.
- Bruns A. Fourier-, Hilbert- and wavelet-based signal analysis: are they really different approaches? *Journal of Neuroscience Methods* 2004; 137: 321–332.
- Bruurmijn MLCM, Peregboom IPL, Vansteensel MJ, Raemaekers MAH, Ramsey NF. Preservation of hand movement representation in the sensorimotor areas of amputees. *Brain* 2017; 140: 3166–3178.
- Cattaneo L, Pavesi G. The facial motor system. *Neuroscience & Biobehavioral Reviews* 2014; 38: 135–159.
- Chin ZY, Ang KK, Guan C. Multiclass voluntary facial expression classification based on Filter Bank Common Spatial Pattern. *Conf Proc IEEE Eng Med Biol Soc* 2008; 2008: 1005–1008.
- Ekman P. The argument and evidence about universals in facial expressions of emotion. In H. Wagner & A. Manstead (Eds.), *Wiley handbooks of psychophysiology*. In: *Handbook of social psychophysiology*. Oxford, England: John Wiley & Sons; 1989. p. 143-164).
- Ekman P, Friesen WV. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. ISHK; 2003.
- Grabski K, Lamalle L, Vilain C, Schwartz J-L, Vallée N, Tropres I, et al. Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Hum. Brain Mapp.* 2012; 33: 2306–2321.
- Herff C, Heger D, de Pestors A, Telaar D, Brunner P, Schalk G, et al. Brain-to-text: decoding spoken phrases from phone representations in the brain. *Front. Neurosci* 2015; 9: 217.
- Iwase M, Ouchi Y, Okada H, Yokoyama C, Nobezawa S, Yoshikawa E, et al. Neural substrates of human facial expression of pleasant emotion induced by comic films: a PET Study. *Neuroimage* 2002; 17: 758–768.
- Kanade T, Cohn JF, Tian Y. Comprehensive database for facial expression analysis. In: *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*. 2000. p. 46–53.
- Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. 2010. p. 94–101.
- Miller KJ. Broadband Spectral Change: Evidence for a Macroscale Correlate of Population Firing Rate? *J. Neurosci.* 2010; 30: 6477–6479.
- Miller N, Maruyama G, Beaver RJ, Valone K. Speed of speech and persuasion. *Journal of Personality and Social Psychology* 1976; 34: 615–624.
- Pineda JA. Sensorimotor cortex as a critical component of an ‘extended’ mirror neuron system: Does it solve the development, correspondence, and control problems in mirroring? *Behavioral and Brain Functions* 2008; 4: 47.
- Ramsey NF, Salari E, Aarnoutse EJ, Vansteensel MJ, Bleichner MG, Freudenburg ZV. Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *Neuroimage* 2018; 180: 301–311.
- Salari E, Freudenburg ZV, Branco MP, Vansteensel MJ, Ramsey NF. Classification of Articulator Movements and Movement Direction from Sensorimotor Cortex Activity. 2018a, submitted article.
- Salari E, Freudenburg ZV, Vansteensel MJ, Ramsey NF. Spatial-Temporal Dynamics of the Sensorimotor Cortex: Sustained and Transient Activity. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 2018b; 26: 1084–1092.
- Schmidt KL, Cohn JF. Human facial expressions as adaptations: Evolutionary questions in facial expression research. *American Journal of Physical Anthropology* 2001; 116: 3–24.
- Smith E, Delargy M. Locked-in syndrome. *BMJ* 2005; 330: 406–409.
- Taner Eskil M, Benli KS. Facial expression recognition based on anatomy. *Computer Vision and Image Understanding* 2014; 119: 1–14.
- Vansteensel MJ, Pels EGM, Bleichner MG, Branco MP, Denison T, Freudenburg ZV, et al. Fully Implanted Brain–

- Computer Interface in a Locked-In Patient with ALS. *New England Journal of Medicine* 2016; 375: 2060–6.
- Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. *Clinical Neurophysiology* 2002; 113: 767–791.
- Wong CH, Birkett J, Byth K, Dexter M, Somerville E, Gill D, et al. Risk factors for complications during intracranial electrode recording in presurgical evaluation of drug resistant partial epilepsy. *Acta Neurochir (Wien)* 2009; 151: 37.





Summary and Discussion

*“The songs are in your eyes, I see them when you smile.
I’ve had enough, I’m not giving up on a miracle drug.”*

Summary and Discussion

What Did We Find?

In the beginning of this thesis we started with two questions, both intended for the development of BCIs. The first question was: how does variability in speech unit pronunciation relate to variability in sensorimotor cortex activity? We looked at three different phenomena that could influence speech related neural activity in the sensorimotor cortex.

In **chapter 1** we demonstrated that some parts of the sensorimotor cortex have a clear relation between vowel production duration and the duration of neural activity, whereas other subareas do not show this relation. We explored for each recording site the temporal dynamics of the neural response. Most areas that are active during speech, showed a peak in neural activity at voice onset. We hypothesized that this may be related to initiation or planning of, for instance, the articulator movements. A second peak at voice offset was visible in some locations and might be related to the articulators moving back to their rest position. In addition to these voice onset and voice offset peaks, in some areas of the SMC, there was clearly sustained neural activity between voice onset and voice offset, the duration of which increased with increasing speech duration. In other areas of the SMC this was not the case. We speculated about the origin of this sustained activity and hypothesized that it might be related to (1) continuous movement or force of muscles, (2) to sustained somatosensory feedback, since a large portion of the electrodes showing sustained responses covered somatosensory areas or (3) to keeping the articulators in their correct position during the pronunciation. Our findings are in agreement with previous single neuron activity research that has indicated that some neurons show only bursts of activity at movement initiation or returning to the rest position, whereas others show activity throughout the static phase of keeping a particular position and even others show a combination of these profiles (e.g. Cheney and Fetz 1980). Since ECoG electrodes record from ensembles of neurons, it can be expected that the proportion of neurons from each of these categories dictates the neural response that is measured by each surface recording site. It remains to be determined, however, what the origins of sustained and transient activity truly are.

In **chapter 2** we demonstrated that repeatedly producing the same sound does not involve the same amount of neural activity with each repetition. Instead, the neural activity amplitude usually decreases with repetitions. These differences in amplitude could not be explained by our measures of how people spoke with each repetition and therefore our results suggest a non-linear relationship between neural activity and motor output. Mainly, there is a decrease of neural activity over the course of repetitions, with the neural activity for the first pronunciation being highest, followed by a non-linear decrease in activity for the following repetitions. In addition, for some subareas, the end of the repetition sequence was associated with a rise in neural activity amplitude, which, as described in chapter 1, may be related to articulators moving back to their rest position. One possible explanation for the changing neural activity associated with the ‘same’ behavior may be that there is some residual variability in articulator movements that we could not measure, such as those of the tongue. Another explanation, however, is that with repeated movements, neurons fire

less or that a smaller number of neurons fire. This phenomenon is called repetition suppression and has been found before for other movements (Hermes et al. 2012; Okada et al. 2018). It has been postulated to be the result of changes in potassium ion influx which hyperpolarizes the cell membrane, causing the neurons to fire less (Grill-Spector et al. 2006). Another explanation that has been put forward, is that in the beginning of a sequence many neurons fire but that only the ones that are task specific or essential continue to fire (Grill-Spector et al. 2006). Future research, in which the movements of all articulators are recorded, may elucidate if the repetition effect is due to articulator movements or if there is indeed a non-linear relation between neural activity and speech movements. Relevant for the development of speech-BCIs would be to investigate if repetition suppression effects are also present during natural speech and not only during isolated repetitions, irrespective of the origin of the effect.

In **chapter 3** we demonstrated that if people pronounce the same sound, the neural activity pattern in the sensorimotor cortex for that specific sound may be different, depending on whether or not another sound was pronounced before and if so, which sound that was. At the same time, the data showed that there is something in the different neural patterns related to one sound that distinguishes them from the patterns of other sounds. This indicates that the different patterns for the same sound share some unique features. Next, we investigated if the existence of multiple activity patterns for the same sound may affect BCI performance and found that it is important to keep these multiple activity patterns in mind when developing and training classifiers. Building a classifier to detect different sounds and training this classifier only on isolated sounds will not lead to the most optimal classification if sounds will be produced in combination with others, for instance. Instead, training a classifier for the multiple patterns that are possible, significantly improved the classification accuracies. Favorably for the development of BCIs, however, is that we found that it is not necessary to have a different model for each sound combination but that a model based on combinations of a certain sound with other sounds can be used to accurately distinguish this sound from other sounds. Since we only focused on the /i/ sound in our studies, the effects on BCI performance of multiple neural activation patterns for other sounds and sound combinations should be further investigated.

In the chapters described above we investigated how variations in speech unit pronunciations and their contexts (i.e. previous pronunciations) are related to variations in sensorimotor cortex activity. The second question that we posed in the beginning was: can variations in movements of individual articulators and facial movements be distinguished from the neural activity patterns in the sensorimotor cortex, and if so, how?

In **chapter 4** we found that movements of different articulators can be accurately distinguished from each other based on ECoG surface recordings in the sensorimotor cortex and we demonstrated that the same is true for different movement directions of the same articulator (i.e. the tongue). Our finding that distinct neural activity is associated with different tongue movement directions is in correspondence to earlier findings of single cell recordings in primates (Murray and Sessle 1992; Yao et al. 2002), but we now show that with surface recordings in humans, similar results can be

obtained. For BCI applications it is important to know where and how to record neural activity, in order to reach high classification performance. We found that the areas needed for classification of both different articulator movements and different movement directions of the same articulator are rather small (approximately 1 cm²). However, we also found that the possible area that *can* be used for classification is larger for classification of different articulators than for different movement directions of the same articulator. Finding the right location for grid placement is therefore crucial and future research needs to establish empirical methods to indicate with non-invasive methods, such as fMRI (see for instance Hermes et al. 2011; Zhang et al. 2013; Siero et al. 2014), which locations promise the best accuracies.

In **chapter 5** we demonstrated that different facial expressions can be distinguished from each other based on sensorimotor cortex activity. We found that both the left and right sensorimotor cortex are good targets for classification of the facial expressions of happy, sad, surprise and disgust. In normal conversation, facial expressions are important and convey much information about somebody's state or intention. We suggest that future communication BCI systems can be enriched by adding, for instance, an emotional avatar to the BCI users' communication output. The reason for this is that with the current BCI applications, emotional nuances are hard to quickly add to a message. Most BCI applications have only focused on the communication of language through the spelling of letters and words, which usually takes much more time than in normal conversation. Typing a response to a joke for instance may take too long, whereas the portrayal of an emotion through a facial expression may be much faster. We showed that facial expressions can be classified, most accurately from the middle aspects of the sensorimotor cortex from areas of approximately 1.40 cm². Future research may extend the current results by investigating more facial expressions, such as for anger and fear.

Implications

The experiments in this thesis address multiple questions as has been outlined above. Here, I will discuss what we have learned from these experiments in more general terms. In the first three chapters, we saw that three sources of variability in the production of elementary speech units result in variability of sensorimotor cortex neural activity. These results indicate that the decoding of speech is a rather complex task. Clearly, the number of different brain activity patterns that a speech BCI has to be able to classify is extensive. We saw, for instance, that a phoneme not only has one representation but that transitions between phonemes result in distinct patterns for the same phoneme, which increases the amount of 'classes'. Additionally, the duration of a phoneme and whether or not it was repeated, influences the neural patterns. Which of these sources of variability has the strongest influence on the classification of natural speech based on neural activity is not yet known but the experiments in this thesis help to understand how the sensorimotor cortex works and how these variances may affect BCI performance. In addition, these experiments may help to indicate what factors to take into account when designing classifiers for a BCI that is based on speech decoding.

I described multiple methods for analyzing the variability in the neural signals and spatial signal patterns, and they may be useful in further investigations to optimize speech BCI decoding.

In chapter 1-2, I developed a method to model the neural response per electrode, which can be used to map duration and repetition effect patterns. In chapter 3 we investigated neural activity levels during sound production in different contexts with a simple ANOVA procedure. In addition to analyzing neural activity dynamics in individual electrodes, in chapter 3-5 I analyzed spatial neural activity patterns, since movements are likely to be represented by a network of neural activity (Schieber 2001), rather than by a single locus. Therefore, neural activity patterns should be analyzed as a whole rather than per electrode. We used template matching classification procedures and found different neural activity patterns for different sound transitions, as well as for different non-speech related articulator/facial movements and tongue movement directions. This suggests that differences in muscle movements are differently represented in the sensorimotor cortex. We show that with template matching methods we can easily and qualitatively investigate differences in neural activity patterns, even for answering fundamental questions.

Finally, the experiments described in this thesis provide insights in where on the sensorimotor cortex information about articulator/facial movements is represented and at what scale (chapter 4 & 5). Hopefully, this may guide us to optimal placement of electrode-grids.

What Can We Learn From Non-BCI Research?

As I indicated above, the number of neural activity patterns that a speech BCI should be able to accurately distinguish makes speech decoding a complex task. Therefore, clever methods need to be developed to tackle the problem of variable signals for the same sounds. Interestingly, similar problems needed to be addressed in the early days of speech recognition based on acoustic signals. Speech recognition underwent a lot of progress over the last decades and possibly, the BCI community may learn from the methods that have been used in speech recognition. One method, for instance, that has increased classification accuracies in speech recognition is the grouping of phonemes in triplets (taking into account the phoneme left and right from the current phoneme), in contrast to looking only at single phonemes or duplets (Lee 1990). Grouping in triplets makes sense, since it is known that both neighboring phonemes (left and right) can have an effect on the articulation of a sound (Davis 2015), but comes with the disadvantage that training a classifier for triplets increases the number of classes due to the large number of possible triplet combinations. One way of tackling this problem, which has been successful in speech recognition, is by smartly grouping the triplets that are similar with respect to their effect on the neighboring sound (Lee 1990). For instance, the /b/ and the /f/ have a similar effect on the right phoneme (example taken from Lee 1990) and triplets containing this phoneme at the beginning may therefore be grouped together. This method increases trainability by decreasing the number of classes while taking into account the variability due to coarticulation effects that may be also beneficial for BCI optimization.

Another issue addressed in speech recognition research, which may be relevant for the BCI community, is the recognition of phonemes produced in function-words such as ‘the’, ‘a’, ‘in’ (Lee 1990). These words are usually very poorly stressed (pronounced) compared to other words and in continuous speech recognition they cause a big portion of the errors (whereas in isolation they may be easily recognized). Training a classifier for the phonemes in these function-words as separate classes can increase continuous speech recognition accuracies. This method points the attention of

a classifier to special cases of phoneme production, meaning that a phoneme may have two classes in a classifier, one for normal words and one for function words. Similar methods may possibly be beneficial for speech BCI performance, although further research into this is needed. There is, however, one important difference between acoustic signal speech recognition and speech BCI, which is the amount of data available to train computer algorithms. Whereas acoustic signals are easily acquired, for recordings of neural signals this is more complex, especially for signals acquired with invasive techniques such as ECoG. This makes standard machine learning solutions hard to implement since they usually require a large amount of training data. Therefore, for the development of effective and smart solutions for neural based speech decoding, it remains crucial to gain a thorough understanding of how the brain, and more specifically the sensorimotor cortex, activates during speech in different contexts. The experiments in this thesis were intended to address exactly this issue and have contributed to increasing our knowledge about the effects of speech variability on the neural signals and have provided methods for analyzing these signals.

Recommendations For Future Speech BCI Research

Although the results in this thesis will help to optimize decoding of speech for BCI usage and help to restore speech in disabled people, more research is needed to accomplish this task since we are still far away from accurate and usable speech based BCIs for locked-in patients. Therefore, I would like to give some recommendations for future speech BCI research, based on what I have learned the last couple of years.

First, based on the experiments in this thesis and the methods that result from them, I propose to create a library of spatiotemporal neural-response-profile-models for speech elements, that can be ‘adjusted’ to fit different contexts. For instance, if we map (similar as in chapter 1) the neural activity duration, we can create models of speech related neural activity, which can be adjusted to how long a phoneme is pronounced, which may potentially increase BCI performance, without the need of speech examples of all possible durations. Similarly, mapping the effects of repetitions (chapter 2) and transitions (chapter 3) into a model may improve speech BCI performance. I have described in this thesis some methods to model neural activity (e.g. in chapter 1 & 2). Although this method gives a simple representation of neural activity, and further improvements need to be made, it has shown to be applicable in multiple situations, for instance in modeling duration effects but also repetition effects. Future research may improve this method for even more accurate modeling.

Second, in the experiments described in this thesis, me and my colleagues observed activity related to speech in large parts of the sensorimotor cortex. For instance, speech-related neural signal changes occurred in areas that are mostly known for controlling *hand* movements. Although most studies so far have focused on the ventral parts of the sensorimotor cortex as a potential target for speech BCI, previous research has also indicated that the hand area is involved during speech movements (Saarinen et al. 2006). One study even was able to accurately classify individual phonemes from the hand knob with needle array recordings (81% accuracy for 9 phonemes and silence; Stavisky et al. 2018). In addition, I performed a pilot study (not described in this thesis) with three participants that had HD-grid electrode coverage of different sizes over the sensorimotor cortex. The goal was to classify 12 Dutch words, and the results showed that two participants who

had a large coverage (128 electrodes), not only on ventral sensorimotor cortex but also up to and including parts of the hand area, showed the highest accuracies (87% and 66% accuracy, 8.33% chance level). In contrast, in one participant who had only coverage over the ventral sensorimotor cortex (64 electrodes) we were not able to classify words with high accuracy (27% accuracy). These findings suggest that inclusion of all informative parts of the sensorimotor cortex is beneficial for speech decoding and these are not limited to the ventral parts.

Third, we found that our HD-grid recordings were able to record locus specific activity at a resolution of as small as 3 mm inter electrode distance (for more discussion about this, see below). This suggests that information can be missed with sparser samplings, for instance with the 10 mm interelectrode distance that is standard for clinical recordings. In addition, previous results on phoneme classification with HD-grid electrodes (72% accuracy for 4 phonemes; Ramsey et al. 2018) and needle arrays (81% accuracy for 9 phonemes; Stavisky et al. 2018) have resulted in high accuracies levels that have thus far not been reported with sparser sampling recordings. These results suggest that dense sampling recordings are beneficial for speech decoding.

Fourth, during speech perception our brain is very well capable of filling in and interpreting missing or distorted words and sounds, based on context information (Plomp 2001). Similarly, speech BCI recordings may also be subject to different forms of distortion, both physiologically and technologically, such as signal variability due to unclear pronunciation or temporal signal connection dropout. To overcome these distortions, in a previous study Herff and colleagues (2015) have added a language corpus to their continuous speech decoder, which resulted in a 75% accurate word prediction for a small set of words. This suggests that it may be beneficial for continuous speech decoding to incorporate context information, such as which words have been said before and which words are likely to follow. Furthermore, a language corpus guiding the prediction of words may even be personalized to the vocabulary of a BCI user, to words that he or she is likely to use. This may be done for instance by analyzing content previously written by the user, such as emails, letters, blogs etc.

Finally, since the data for invasive neural recordings is limited, especially within one research lab, I think it is essential that, as a BCI community, we need to agree on which experiments can bring the field forward and join forces in the collection of data and sharing it. Furthermore, the task of speech decoding involves many different aspects and requires many different forms of expertise. Collaborations between labs and taking advantage of each of their expertises, is therefore crucial if we want to increase the amount of available data and gain any major progress. In addition, data need to be collected over longer periods of time. Currently, most ECoG based speech BCI research is done with epileptic patients that are in the hospital for usually not more than a week. How classifiers perform over longer periods of time is therefore not well investigated. Furthermore, it is generally not possible in these clinical settings to analyze the data, adjust the classifier and repeat the experiment with the same participant. In other words, it is difficult to try out what works and what not, for optimal speech decoding. A previous study, however, has placed electrodes over a year in a patient with LIS and this has impressively gained insights in how to make use of neural signals for BCI usage (Vansteensel et al. 2016). In addition, a similar experimental setting as described by Vansteensel and colleagues (2016) would also be very informative for speech BCI research.

Limitations & Additional Future Directions

In addition to the future directions described in the previous sections, there are some limitations and additional future directions that I will discuss here. First, although we have investigated some sources of variability in the production of elementary speech units and their relation to the variability in sensorimotor cortex activity, for most of them it is still not clear how they may affect BCI performance. For instance, it is still unclear how differences in phoneme duration and repetition affects BCIs performance in real-time natural speech classification. The same is true for transitions between sounds, since we only investigated a limited number of sound transitions. Future research may focus on the question of how these sources of variability affect BCI performance in natural speech settings. Besides, we only investigated three sources of variability, and other sources may exist, such as for instance volume or pitch and intonation.

Second, since the number of patients that undergo ECoG measurements is small, in most of our experiments only a limited number of participants could be included. Therefore, it is hard to say how well the results generalize to larger groups of subjects. For instance, it remains unclear if the anatomical locations that show a relation between neural activity duration and speech duration are similar over subjects. The same holds for areas that show repetition effects or coarticulation effects. However, the fact that we find these effects consistently in subjects (even in such a small research population), indicates that these effects are likely to be there for most people.

Another limitation is that in some subjects we could not record with high-density ECoG grids. However, since we used both clinical ECoG grids (inter electrode distance: 10 mm) and high-density ECoG grids (inter electrode distance: 3 or 4 mm) we were able to compare the two grid types with respect to how much information about the neural activity they deliver. We found more and different neural patterns with the HD-grids than with the standard grids. Importantly, neighboring electrodes in the HD-grids sometimes showed different neural responses, which we would have missed with standard grid recordings. There seems to be much detail in the neural activity patterns that can only be uncovered with high resolution sampling. Furthermore, since the electrodes of HD-grids are smaller in diameter they may be more specific (Camuñas-Mesa and Quiroga 2013). Indeed, large populations of neurons do not necessarily behave in a similar way. Therefore, data sampled from a large population of neurons is more likely to represent a mixture of local neural processes than sampling from only a small population. Although the relation between electrode size and signal stability needs to be investigated further, as well as their influence on BCI performance stability, we believe that the current HD-grids provide more information than standard grids and can be beneficial for speech-based BCI performance.

Finally, since we explored the relation between sensorimotor cortex activity and overt speech, as well as overt articulator and facial movements, it would be interesting to see if the results that we find are similar for attempted movements. Currently, investigating attempted speech and facial movements is rather complicated with ECoG due to the limited number of paralyzed subjects that can be included for studies such as the ones we did. Studies with BCI target users have mostly focused on other areas than speech, and it will be exiting for the future to move more attention to decoding attempted speech. With the availability of more and better BCI applications and of better recording equipment, we will be able to do more research with the target population of BCI and hopefully answer these questions in the future.

References

- Camuñas-Mesa LA, Quiroga RQ (2013) A Detailed and Fast Model of Extracellular Recordings. *Neural Comput* 25:1191–1212. doi: 10.1162/NECO_a_00433
- Cheney PD, Fetz EE (1980) Functional classes of primate corticomotoneuronal cells and their relation to active force. *J Neurophysiol* 44:773–791
- Davis J (ed) (2015) *Phonetics: fundamentals, potential applications and role in communicative disorders*. Nova Publishers, New York
- Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14–23. doi: 10.1016/j.tics.2005.11.006
- Hermes D, Vansteensel MJ, Albers AM, et al (2011) Functional MRI-based identification of brain areas involved in motor imagery for implantable brain-computer interfaces. *J Neural Eng* 8:025007. doi: 10.1088/1741-2560/8/2/025007
- Hermes D, Siero JCW, Aarnoutse EJ, et al (2012) Dissociation between Neuronal Activity in Sensorimotor Cortex and Hand Movement Revealed as a Function of Movement Rate. *J Neurosci* 32:9736–9744. doi: 10.1523/JNEUROSCI.0357-12.2012
- Lee KF (1990) Context-independent phonetic hidden Markov models for speaker-independent continuous speech recognition. *IEEE Trans Acoust Speech Signal Process* 38:599–609. doi: 10.1109/29.52701
- Murray GM, Sessle BJ (1992) Functional properties of single neurons in the face primary motor cortex of the primate. III. Relations with different directions of trained tongue protrusion. *J Neurophysiol* 67:775–785. doi: 10.1152/jn.1992.67.3.775
- Okada K, Matchin W, Hickok G (2018) Phonological Feature Repetition Suppression in the Left Inferior Frontal Gyrus. *J Cogn Neurosci* 1–9. doi: 10.1162/jocn_a_01287
- Plomp R (2001) *The Intelligent Ear: On the Nature of Sound Perception*. Psychology Press
- Ramsey NF, Salari E, Aarnoutse EJ, et al (2018) Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage*. doi: 10.1016/j.neuroimage.2017.10.011
- Saarinén T, Laaksonen H, Parviainen T, Salmelin R (2006) Motor Cortex Dynamics in Visuomotor Production of Speech and Non-speech Mouth Movements. *Cereb Cortex* 16:212–222. doi: 10.1093/cercor/bhi099
- Schieber MH (2001) Constraints on Somatotopic Organization in the Primary Motor Cortex. *J Neurophysiol* 86:2125–2143
- Siero JC, Hermes D, Hoogduin H, et al (2014) BOLD matches neuronal activity at the mm scale: A combined 7 T fMRI and ECoG study in human sensorimotor cortex. *NeuroImage* 101:177–184. doi: 10.1016/j.neuroimage.2014.07.002
- Stavisky S, Rezaei P, Willett F, et al (2018) Decoding Speech from Intracortical Multielectrode Arrays in Dorsal “Arm/Hand Areas” of Human Motor Cortex. 5
- Vansteensel MJ, Pels EGM, Bleichner MG, et al (2016) Fully Implanted Brain–Computer Interface in a Locked-In Patient with ALS. *N Engl J Med* 375:2060–6. doi: 10.1056/NEJMoa1608085
- Yao D, Yamamura K, Narita N, et al (2002) Neuronal Activity Patterns in Primate Primary Motor Cortex Related to Trained or Semiautomatic Jaw and Tongue Movements. *J Neurophysiol* 87:2531–2541. doi: 10.1152/jn.2002.87.5.2531
- Zhang D, Song H, Xu R, et al (2013) Toward a minimally invasive brain–computer interface using a single subdural channel: A visual speller study. *NeuroImage* 71:30–41. doi: 10.1016/j.neuroimage.2012.12.069





Conclusion

Conclusion

In this thesis, we showed how three sources of variability in phoneme production are related to variability in sensorimotor cortex activity and we discussed how this may affect BCI performance. Furthermore, we showed how different articulator movements and movement directions, as well as facial movements, lead to unique and distinct sensorimotor cortex neural activity patterns, which can be used to improve communication BCIs. We show that that these patterns can be quite detailed since different movements can be distinguished from small parts of the SMC. This research is a step into the direction of better speech and communication aids for paralyzed subjects by means of BCIs. I hope that this work will have an impact on the future directions of BCI research.



Nederlandse Samenvatting

Nederlandse Samenvatting

Wat is een BCI en waar wordt het voor gebruikt?

Mensen communiceren op vele verschillende manieren, bijvoorbeeld door middel van spraak, maar ook wanneer ze een brief schrijven, een sms-bericht sturen of iemand iets duidelijk proberen te maken met een gebaar of een gezichtsuitdrukking. Voor mensen die geheel of bijna geheel verlamd zijn is het echter een stuk ingewikkelder om zichzelf uit te drukken. Dit komt omdat, hoewel we er eigenlijk bijna nooit bij stil staan, voor al deze vormen van communicatie bewegingen nodig zijn die weg vallen wanneer iemand geheel of bijna geheel verlamd raakt. Mensen die (vrijwel) geheel verlamd zijn, maar nog wel kunnen denken en voelen lijden aan het zogeheten Locked-In Syndroom (LIS) en zijn, als het ware, gevangen in hun eigen lichaam. Nieuwe technologieën hebben zich in de laatste jaren echter ontwikkeld om deze mensen te helpen met communiceren. De meeste van deze technologieën maken daarbij gebruik van de laatste overgebleven vormen van bewegingen die deze patiënten nog wel kunnen maken, zoals bijvoorbeeld oogbewegingen. Omdat de ogen veelal nog wel bewogen kunnen worden, kunnen deze gebruikt worden om te communiceren. Zo kan een verzorger of familielid de patiënt een vraag stellen waarbij bijvoorbeeld een oogbeweging omhoog 'ja' betekent en een beweging naar beneden 'nee'. Ook kan bijvoorbeeld een eye-tracker gebruikt worden. Met dit apparaat kan gemeten worden waar iemand naar kijkt op een scherm. Als dit scherm gevuld is met letters, kan dit apparaat bepalen welke letter iemand wil spellen, en zo kan een hele zin uitgeschreven worden.

Er zijn echter ook mensen voor wie zelfs deze laatste vormen van bewegingen uiteindelijk verdwijnen, bijvoorbeeld in sommige gevallen van Amyotrofische Laterale Sclerose (ALS). Dit is een ziekte waarbij steeds meer motorische functies uitvallen doordat neuronen die hiervoor verantwoordelijk zijn afsterven. Gelukkig is er voor deze mensen een systeem ontwikkeld welke niet afhankelijk is van enige vorm van beweging. Zo'n systeem wordt een brain-computer interface (BCI) genoemd en meet hersensignalen welke de gebruiker heeft geleerd te controleren. Deze signalen kunnen vervolgens door een computer geïnterpreteerd worden en omgezet naar bijvoorbeeld uitgeschreven tekst of gesynthetiseerde spraak. Zo kan iemand die geheel verlamd is nog steeds communiceren.

Spraak BCIs

Hoewel er vele vormen van zo'n BCI systeem bestaan, wordt er de laatste jaren veel onderzoek gedaan naar de ontwikkeling van een systeem waarbij LIS patiënten letters of woorden zouden kunnen denken of proberen te zeggen, en welke de BCI vervolgens herkent op basis van het hersensignaal. De hersengebieden die hiervoor het best kunnen worden gebruikt zijn de primaire motorische en somatosensorische gebieden (samen de sensomotorische cortex). Deze gebieden houden zich voornamelijk bezig met het aansturen van bewegingen en het verwerken van informatie over de positie van-, en gevoel in-, het lichaam. Hoewel onderzoek heeft aangetoond dat spraak-BCI systemen potentie hebben om in de toekomst gebruikt te gaan worden is er tot op heden geen systeem ontwikkeld dat robuust en accuraat genoeg is om thuis door patiënten gebruikt te worden. Een mogelijke reden hiervoor is dat de hersensignalen die betrokken zijn bij de aansturing van spraakbewegingen (de signalen die gebruikt worden voor spraak BCI) erg variabel zijn. Zo

kunnen bijvoorbeeld de hersensignalen die nodig zijn om een klank uit te spreken beïnvloed worden door de manier waarop je iets zegt, of door de klanken die je ervoor of erna zegt. Welke factoren een invloed hebben en hoe die invloed zich manifesteert is echter nog niet geheel duidelijk. Daarom heb ik in deze thesis een drietal fenomenen onderzocht welke eventueel een invloed zouden kunnen hebben op spraak-gerelateerde hersensignalen en ik heb een methode ontwikkeld om deze fenomenen beter in kaart te kunnen brengen.

In **hoofdstuk 1** toon ik aan dat sommige delen van de sensomotorische cortex net zo lang actief zijn als dat iemand een bepaalde klank uitspreekt. Andere gebieden laten zo'n relatie echter niet zien en zijn alleen aan het begin en/of aan het einde van een klank actief. Duidelijk is dat de hersenactivatie afhankelijk kan zijn van hoe lang iemand praat en we speculeren in dit hoofdstuk over de mogelijke oorzaken hiervan. Het kan bijvoorbeeld zijn dat de continue hersenactiviteit in sommige gebieden gerelateerd is aan (1) een continue beweging of kracht van spieren, die optreedt tijdens het uitspreken van een klank, of dat (2) er continue somatosensorische terugkoppeling (feedback) plaatsvindt, aangezien een groot deel van de gebieden die een continu hersensignaal vertonen zich bevinden in de primaire somatosensorische cortex, of dat (3) deze hersenactiviteit gerelateerd is aan de positie van sommige articulatoren welke gedurende het uitspreken van een klank op dezelfde positie moeten blijven. De gebieden die niet continu actief zijn maar alleen aan het begin en/of einde van een klank, hebben wellicht een rol in de initiatie van een beweging of de motorische planning die daarvoor nodig is. Dit kan ook verklaren waarom sommige gebieden een (tweede) piek in activiteit laten zien aan het einde van de klank, omdat er dan weer een teruggaande beweging wordt ingezet.

In **hoofdstuk 2** toon ik aan dat het herhaaldelijk uitspreken van dezelfde klank niet gepaard gaat met telkens dezelfde hersenactiviteit. De hoogte van de neurale activiteit gaat namelijk meestal non-lineair omlaag terwijl de uitspraak ongeveer hetzelfde blijft. Een verklaring voor de afname in neurale activiteit voor 'eenzelfde' beweging zou kunnen zijn dat we niet alle bewegingen hebben kunnen meten (bijvoorbeeld die van de tong) en dat deze bewegingen afnemen over repetities. Echter, een andere verklaring zou kunnen zijn dat bij herhaalde bewegingen, de neuronen minder gaan vuren of dat er minder neuronen gaan vuren. Dit fenomeen heet 'repetitie onderdrukking' (repetition suppression) en kan worden veroorzaakt doordat bij herhaalde beweging de kalium toevoer naar de neuronen verandert waardoor de neuronen minder 'gevoelig' worden en minder gaan vuren. Een andere mogelijkheid is dat aan het begin van een bewegings-sequentie veel neuronen vuren en dat na verloop van tijd alleen de essentiële neuronen blijven vuren, waardoor het signaal afzwakt. Ondanks de afname in neurale activiteit over repetities ging in sommige gebieden de activiteit weer omhoog aan het einde van een reeks herhalingen. De toename in activiteit aan het einde van een beweging zagen we ook al in hoofdstuk 1, en zou eventueel te maken kunnen hebben met de articulatoren die weer terug bewegen naar hun rust positie. Hierbij wordt er een nieuwe/andere beweging ingezet welke nieuwe neurale activiteit zou kunnen veroorzaken.

In **hoofdstuk 3** toon ik aan dat als mensen een klank uitspreken, de neurale activiteit voor die klank afhankelijk is van welke klank ervoor is uitgesproken. Voor één klank zijn er dus eigenlijk meerdere hersenpatronen mogelijk. Echter, de verschillende neurale patronen voor één klank blijken wel meer op elkaar te lijken dan die van andere klanken. Daarom onderzochten we verder

hoe de verschillende patronen voor dezelfde klank gegroepeerd kunnen worden en het best onderscheiden kunnen worden van de neurale patronen van andere klanken. We vonden dat wanneer je een BCI maakt die klanken van elkaar probeert te onderscheiden, het belangrijk is om rekening te houden met de meerdere neurale patronen die mogelijk zijn voor één klank maar dat het niet nodig is om voor iedere klankcombinatie een apart model te maken. Één model welke gebaseerd is op het gemiddelde van alle combinaties kan goed voorspellen welke klank iemand zegt. Dit is handig voor BCIs die zich op spraak richten, aangezien er veel klankcombinaties mogelijk zijn en een classificatie algoritme erg complex zou worden als het rekening moest houden met al deze combinaties. Echter is het wel essentieel om tijdens het trainen van een algoritme, dit algoritme voldoende voorbeelden van verschillende combinaties aan te bieden.

In de bovenstaande hoofdstukken beschreef ik hoe variaties in spraak -en hoe voorgaande klanken- een invloed hebben op de hersenactiviteit in de sensomotorische cortex. Echter, deze metingen zijn in een experimentele setting uitgevoerd en welke van deze fenomenen ook een invloed hebben op de hersenactiviteit tijdens natuurlijke spraak zal nog moeten blijken. Duidelijk is wel dat er variatie bestaat in de hersensignalen voor dezelfde klank en dat dit invloed kan hebben op de prestaties van een BCI. De methoden die ontwikkeld zijn om de bovenstaande vragen te beantwoorden, en welke beschreven staan in deze thesis, kunnen hopelijk een bijdragen leveren aan verder onderzoek naar spraak-gerelateerde hersenactiviteit en het ontwikkelen van nog betere BCIs.

Articulator bewegingen en gezichtsuitdrukkingen

Naast spraakgerelateerde signalen heb ik ook niet-spraakgerelateerde hersensignalen onderzocht van bijvoorbeeld geïsoleerde tong, lip of larynx bewegingen maar ook van gesynchroniseerde bewegingen van de mond, neus en ogen tijdens het maken van gezichtsuitdrukkingen. De reden hiervoor is dat we een beter beeld wilden krijgen van hoe losse articulator bewegingen gerepresenteerd zijn in het brein maar ook of niet-spraakgerelateerde bewegingen zoals die van gezichtsuitdrukkingen gebruikt zouden kunnen worden om een BCI te optimaliseren.

In **hoofdstuk 4** toon ik aan dat bewegingen van verschillende articulatoren (de lippen, tong, kaak en larynx) van elkaar kunnen worden onderscheiden op basis van de signalen van elektroden die op de sensomotorische cortex geplaatst worden. Ook bewegingen van dezelfde articulator, namelijk de tong, in verschillende richtingen kunnen van elkaar onderscheiden worden. Verder is het voor invasieve BCI-applicaties van belang om zoveel mogelijk informatie te halen uit een zo klein mogelijk gebied. Dit in verband met het minimaliseren van risico's die gepaard gaan met het chirurgisch plaatsen van elektroden op het brein onder de schedel. In dit hoofdstuk laten we zien dat voor de classificatie van verschillende articulatoren en van verschillende bewegingsrichtingen van de tong, een relatief klein gebied nodig is (ca. 1 cm²). Echter, het gebied dat gebruikt zou kunnen worden voor het classificeren van verschillende articulatoren is wel groter dan voor het classificeren van verschillende richtingen van de tong.

In **hoofdstuk 5** toon ik aan dat verschillende gezichtsuitdrukkingen (blij, verdrietig, verbaasd en walging) van elkaar kunnen worden onderscheiden op basis van sensomotorische hersenactiviteit. Dit kan zowel van de linker- als van de rechterhersen helft van een gebied van ongeveer 1.40 cm². De mogelijkheid om op een snelle manier emoties te tonen door middel van

een BCI zou voor patiënten een belangrijke toevoeging kunnen zijn tijdens communicatie met anderen omdat gezichtsuitdrukkingen veel informatie bevatten over iemands emotie en intentie. Voorsnog is de mogelijkheid om emoties tegelijk met een getypt bericht uit te drukken nog niet geïmplementeerd met de huidige BCI systemen maar dit onderzoek moedigt vervolgonderzoek hiernaar erg aan.

In de laatste twee hoofdstukken laat ik zien dat geïsoleerde articulator bewegingen maar ook gezichtsuitdrukkingen een additionele bron van BCI controlesignalen zou kunnen zijn voor communicatie BCIs. Verder kunnen deze experimenten gebruikt worden om meer te leren over welke gebieden het meest informatief zijn voor zulke applicaties evenals hoe groot het gebied moet zijn waarvan gemeten moet worden om de beste prestaties te verkrijgen.

Conclusie

In deze thesis laat ik zien hoe drie bronnen van variatie in het uitspreken van klanken gerelateerd zijn aan variatie in hersensignalen. Verder bespreek ik hoe dit de prestaties van BCIs kan beïnvloeden. Ook laat ik zien dat verschillende articulator bewegingen en bewegingsrichtingen, maar ook verschillende gezichtsuitdrukkingen, een uniek activiteitspatroon hebben in de sensomotorische cortex. Deze patronen kunnen gebruikt worden om BCIs te verbeteren en te verrijken met meer controlemogelijkheden. Daarnaast laat ik zien dat deze patronen erg gedetailleerd kunnen zijn, aangezien deze bewegingen kunnen worden onderscheiden van slechts een klein gebied.

Dit onderzoek is een stap in de richting van betere spraak- en communicatiehulpmiddelen door middel van een BCI voor mensen die verlamd zijn. Ik hoop dat dit een impact zal hebben op vervolgonderzoek naar communicatie BCIs.



Acknowledgements

Dankwoord

Acknowledgments / Dankwoord

I would like to thank many people, here goes...
Ik wil graag veel mensen bedanken, hier gaat 'ie dan...

The participants/ De participanten

First of all, I would like to thank all people that participated in my experiments for their incredible devotion, especially in the case of the epilepsy patients, who contributed to science while undergoing difficult and scary times in their preparation for brain surgery. Without them I would not have been able to create this thesis.

Allereerst wil ik graag alle mensen bedanken die hebben mee gedaan aan mijn experimenten voor hun tomeloze inzet, met name de epilepsiepatiënten, welke hun bijdrage hebben geleverd aan de wetenschap terwijl ze door een moeilijke en soms beangstigende tijd moesten gaan in de voorbereiding op hun hersenoperatie. Zonder hen zou ik deze thesis nooit gemaakt kunnen hebben.

My Promotors

Zac Freudenburg, I don't think there are many co-promotors like you. To be honest, you were more like one of 'the guys'. We shared many hotel rooms together during our trips to San Diego, Washington, San Francisco, Freiburg and we talked about music many times. You sing in our band and you ran us to second place in the UMC relay race more than once. Oh yeah, and if I needed scientific input, you gave me more than I could handle. If I was frustrated with my work or didn't know how to continue, I just needed to ask you for advice. Besides, you were always right. If I wanted to go my own stubborn way, in the end I had to conclude you were the one that had always been right from the start (even though you've let me go my own way). I think indeed you were right, in contrary to meat loaf, that *two out of three* might be bad!

Mariska Van Steensel, thank you so much for being so thorough with all my papers, abstracts, posters, emails or anything whatsoever that I had to write, make or create. You picked out all the tiniest mistakes that I made which I'm really grateful for. If whatever I was doing passed you, I could be sure that it was correct and well written. You improved my writing so enormously, I cannot even put it into words. The devotion and speed with which you worked and supervised me (and so many others) really impressed me, as well as your true belief in people! Thanks for being my co-promotor.

Nick Ramsey, thanks for giving me the opportunity to learn so much, not only the last four years as a PhD student but also already before as an intern. You were the one who sent me to the USA, which was for sure one of the best periods in my life, where I learned ECoG analysis. You were also the one who, for some strange reason, had the courage to take me back and you made me look at brain signals, aka squiggly lines, for four years! I admire your effort to share data and knowledge with other labs and I hope this will continue and grow in the future. The decoding of speech is a

frontier that you started to cross and even after many years still are devoted for to do so, which is laudable.

My colleagues

Thanks to all my other colleagues;

Anna Gaglianese for talking about Dutch people and Italian people and how wonderfully weird they all are. **Annemiek Elberste**, for taking care of Nick (and all my administration). And for talking with me about non-scientific stuff, I enjoyed dropping by. **Anouk Smits**, for your statistical knowledge and your humor. **Benny van der Vijgh**, for understanding and acknowledging the man-hours I spent working on Xelo. **Dora Hermes**, for your continuous devotion to help me with my work and future goals. **Elmar Pels**, for convincing me to go to Iceland and for your artistic view on pie-charts. But also, for ringing up patients and convincing them to get HD grids, delivering me with great data and the group with extra IEMU shifts. **Erik Aarnoutse**, for taking care of all technical matters and for (usually) leaving an extra piece of pie for the rest of the group. **Francisco Guerreiro Fernandes**, for baking my apple-raspberry pie and bringing it to the lab (too bad I missed it!). **Gio Piantoni**, for teaching me that ‘Gaussian’ is with two s’es and for helping me setting up Python. **Julia Berezutskaya**, for driving to new lab-collaborations (literally 6 hours to Freiburg). **Mariana Pedroso Branco**, for taking so much good care of me. You looked after me, better than I did myself sometimes. You were always there for me and I’m grateful to call you my friend. Without your work I wouldn’t have had all the electrode coordinates of the subjects in my studies. But most importantly, you have a special place in my heart and there are just too many things I would need to thank you for. **Mark Bruurmijn**, for your interesting view on the world. There were many times in which I felt delighted to discuss with you the important nonsense of life. Thanks also for explaining all the difficult scientific subjects in such simple ways that even I could understand it! **Martijn Jansma**, for your view on statistics science. **Mathijs Raemaekers**, for taking the time to help me on the really difficult fMRI related questions. **Max van den Boom**, for helping me on the fMRI data and giving it back to me just before we discovered it was useless. And also, for the many drinks and cocktails we had at conferences and summer schools. **Meron Vermaas**, for playing trombone in one of my songs (and of course in the Gamma Band!). **Miek de Dreu**, for asking me to join for coffee all the time and your pleasant presence. **Philippe Cornelisse**, for making sure the MRI scanner was doing what it supposed to do (well most of the time then...) and for playing bass or guitar, whichever we were in need of. Thanks also for your literature study on facial expressions. **Sacha Leinders**, for giving me a great record of Rod Stewart and for patiently waiting on me to fix Xelo. **Sander van Dieren**, for great and interesting conversations about music and the environment and for sometimes understanding me better than I understood myself. **Wouter Schellekens**, for wearing pink sunglasses.

My office-mates

To **Mark, Max and Elmar**: et bene pendentēs!

Students

I would like to thank the master students I had the pleasure to supervise. They did some great and interesting experiments and I loved working with them; **Andreas Wolters, Caroline de Booij, Joosje Kist & Stefan Pullen.**

Colleagues outside of the lab

There are also some people in the UMC outside of our lab that I'm grateful to; **Nathalia Petridou & Jeroen Siero**, for developing the Multiband EPI sequences. The whole KNF-team, especially **Cyrille Ferrier, Frans Leijten, Geertjan Huiskamp, Tineke Gebbink**, and the neurosurgery team, **Kai Miller, Peter Gosselaar, Peter van Rijen, Pieter van Eijsden** for their care of the patients and letting us work with them.

Also, I would like to thank **Etske Ooijevaar**, for teaching me to record and analyze ultrasound data and for letting me borrow the equipment. **Dick Hermes**, for teaching me the ins and outs of speech recognition and for providing me with vowel onset detection algorithms. **Christian Herff & Emily Mugler** for the fun and inspiring conferences and data collaborations.

The Gamma Band

A.k.a. Groove the least, a.k.a. Fourier and the magnificent sinusoids, for the joyful evenings of making awesomely bad music.

Previous mentors

Eric Leudthardt, and his team, for teaching me how to analyze ECoG data and for the amazing time in St. Louis. **Martin Bleichner**, for having the incredible patience to teach me how to write a paper. If it wasn't for you going back and forth a zillion times on my first report, I'm not sure if I would ever have been able to write proper science articles. **Stella Donker**, for your belief in me even after all these years, for the coffee and for teaching me to appreciate the beauty in simple science.

Mijn familie

Mijn ouders, **Ria & Bram Salari**, omdat ze zulke geweldige ouders zijn en voor hun rotsvaste vertrouwen in mij. Jullie hebben mij altijd mijn eigen ding laten doen en alles in het werk gesteld om dat mogelijk te maken. Ik kan me geen betere jeugd hebben gewenst.

Anoek Wijnen, omdat je familie voor me was gedurende alle jaren van mijn PhD. Jij stond achter me als het zwaar was en maakte dat ik dingen durfde die ik anders niet gedurfd had!

Opa en oma, gewoon omdat ze speciaal zijn.
En natuurlijk de rest van mijn familie!

My friends/ Mijn vrienden

Some of which dared to participate in some of my experiments!

Waarvan sommige het zelfs aandurfdn om met een experiment van mij mee te doen!

Frans, Imre, Jorrit, Mark, voor jullie vriendschap al sinds de basisschool. **Camiel, Jeroen, Wannes**, voor jullie vriendschap al sinds de middelbare school en voor de muziek. **Ingrid, Joris, Nathalie, Patrick**, voor jullie vriendschap al sinds de middelbare school. **Floris, Henk, Lisette, Maartje, Sergio, Tobias, Welmoed, Xander**, voor jullie vriendschap, het gezellige wonen en muziek maken. **Eline, Nicole**, voor jullie vriendschap al sinds de Master. **Beatrice, Cristina, Felipe, Flora, Mariantonia, Natsuki**, for your friendship and our amazing adventures in St. Louis. **Lenny, Mrinal, Ravi, Nick, David**, for your friendship and for taking care of me in St. Louis.



Scientific Achievements

Scientific Achievements

Journal Articles

Bleichner M. G., Jansma J. M., **Salari E.**, Freudenburg Z. V., Raemaekers M., Ramsey N. F. (2015). "Classification of Mouth Movements Using 7 T fMRI". *Journal of Neural Engineering* 12, nr. 6: 066026.

Ramsey N. F., **Salari E.**, Aarnoutse E. J., Vansteensel M. J., Bleichner M. G., Freudenburg Z. V. (2018). "Decoding Spoken Phonemes from Sensorimotor Cortex with High-Density ECoG Grids". *NeuroImage* 180, nr. Pt A: 301–11.

Salari, E., Freudenburg Z. V., Vansteensel M. J., Ramsey N. F. (2018) "Spatial-Temporal Dynamics of the Sensorimotor Cortex: Sustained and Transient Activity". *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26, nr. 5: 1084–92.

Salari, E., Freudenburg Z. V., Vansteensel M. J., Ramsey N. F. (2018) "Repeated Vowel Production Affects Features of Neural Activity in Sensorimotor Cortex". *Brain Topography*

Salari, E., Freudenburg Z. V., Vansteensel M. J., Ramsey N. F. (2018) "The influence of prior pronunciations on sensorimotor cortex activity patterns during vowel production". *Journal of Neural Engineering* 15, nr. 6: 066025

Conference proceedings

Salari E., Freudenburg Z.V., Vansteensel M.J., Ramsey N.F., Temporal dynamics of mouth motor cortex activity during Speech. Proceedings of the 6th International Brain-Computer Interface Meeting, organized by the BCI Society, 2016, pp. 214, doi: 10.3217/978-3-85125-467-9-214

Conference abstracts

Salari E., Freudenburg Z.V., Vansteensel M.J., Ramsey N.F., Unravelling temporal dynamics of sensorimotor cortex activity during speech for bci decoding. Society for Neuroscience, San Diego, USA, 2016

Salari E., Freudenburg Z.V., Vansteensel M.J., Ramsey N.F., Speech movement parameters reflected in sensorimotor cortex activity. Society for Neuroscience, Washington D.C., USA, 2017

Salari E., Freudenburg Z.V., Vansteensel M.J., Ramsey N.F, Towards decoding speech: effects of prior phonemes on sensorimotor cortex activity during sequential vowel production. 7th International Brain-Computer Interface Meeting, Pacific Grove., USA, 2018

Invited talks

Salari E. (2017). Guest speaker at the Symposium on Speech Decoding from 24/7 ECoG Recordings, Freiburg, Germany.

Salari E. (2018). Guest speaker at the workshop “Progress in Decoding Speech processes using intracranial signals” during the 7th International Brain-Computer Interface Meeting, Pacific Grove, USA.



About the Author

About the Author

Efraïm Salari was born on March 20th of 1991. He grew up in the little town of Culemborg in the middle of the Netherlands where he spent his elementary and high school period. After this he finished his bachelor psychology cum laude at the Utrecht University with a specialization in neurobiological psychology. In addition, he followed the Von Humboldt social-sciences-honors-track. He continued his education with a master in neuroscience and cognition at the Utrecht University, during which he worked as an intern in the lab of Prof. dr. Nick Ramsey. His research concerned the decoding of speech and non-speech mouth movements using fMRI signals. Afterwards he moved to the USA to start an internship at the Washington University in Saint Louis. Here he worked in the lab of Prof. dr. Eric Leuthardt where he investigated the relationship between intelligence and brain networks using ECoG signals. In addition, he wrote a thesis about the contemporary theories on the neuroscientific explanation of consciousness. After moving back to Utrecht, he spent the last 4 years as a PhD candidate in the lab of Prof. dr. Nick Ramsey, investigating the relation between speech-, articulator-, and facial movements and the variability in neural signals in the sensorimotor cortex. The goal of this research was to optimize the decodability of speech from neural signals for the development of speech BCIs, as well as adding non-speech related movements such as those of facial expressions for BCI control. During these experiments he developed methods for analyzing and modeling speech-related neural activity, which may help develop smarter communication BCIs in the future.