Università degli Studi di Ferrara

DOCTORAL COURSE IN

TRANSLATIONAL NEUROSCIENCE AND NEUROTECHNOLOGIES

CYCLE XXXII

COORDINATOR Prof. Luciano Fadiga

Exploring micro-Electrocorticographic Signals: from Animal Models to Humans

Scientific/Disciplinary Sector (SDS) BIO/09

Candidate

Dott. Delfino Emanuela

Supervisor

Prof. D'Ausilio Alessandro

Co-Supervisor

Prof. Fadiga Luciano

Year 2016/2019

Abstract

Introduction. Electrocorticography (ECoG), a neural recording technique employed in both clinical and research applications, is characterized by a relatively high spatiotemporal resolution. ECoG has extremely low susceptibility to noise and motion artefacts compared to other techniques, such as electroencephalography (EEG). Recently, the possibility of decoding speech from ECoG signals has been investigated with promising results, significantly advancing the clinical viability of using speech-related brain-computer interfaces (BCI) to restore communication. Speech neuroprosthetic devices aim to improve the quality of life of people suffering from communication deficits because of serious motor disabilities. In such patients, vocalization might not be possible due to severe paralysis, even though language areas are still intact. However, two technical aspects shall be improved before researchers can start clinical trials in patient populations. The first key improvement is concerning the tolerability of chronic ECoG implants. Standard ECoG grids cover different brain areas recruited in language processing, which is an advantage for speech decoding. However, this requires invasive procedures due to the large size of the grid, while its stiffness can lead to inflammatory responses. One critical improvement could involve flexible high-density micro-grids directly placed over eloquent areas for speech production. This approach would increase the spatial resolution by maximizing the specificity of the recorded signals; furthermore, it would minimize the risk of damaging the cortex. The second key improvement of the current approaches extends beyond the technical implantation limits. To make use of promising results obtained in speech decoding from neuronal signals for neuroprosthetic applications, more attention should be paid to the feasibility of their use in a natural setting, such as in communication deficits. One critical issue in the development of assistive devices is the lack of detectable speech-related events to control the decoding. Detecting speech-related motor intentions would represent a fundamental step toward improving speech neuroprosthetics. Furthermore, this achievement could function as a trigger to initiate decoding whenever an explicit alignment is not possible (e.g., the case of covert speech). Since a vocal cue is

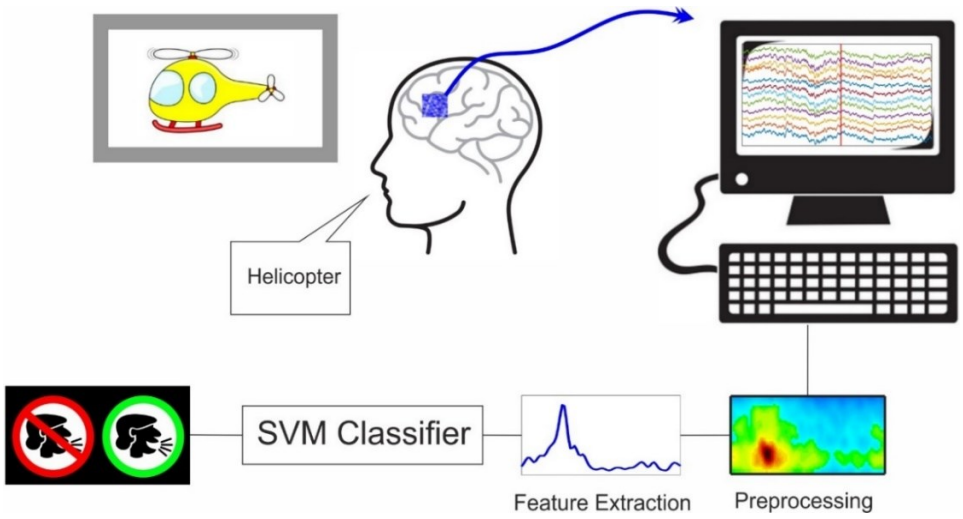
employed to start the most common virtual assistants (e.g., Google Assistant, Alexa, Siri), a neuronal cue to activate the speech decoder is fundamental in applications for patients unable to speak.

Methods. First, a new generation of devices known as micro-ECoG (µECoG, electrode pitch below 1 mm) arrays was tested in rats to determine the best recording configuration in terms of reference and ground connections: *Single-Ended Screw* (without reference), *Differential or Single-Ended Reference* (with reference). Several experiments were conducted for the setup validation using adult Long Evans rats. The ECoG devices were placed over the barrel cortex, and signals were recorded during mechanical stimulation of the whiskers. Next, two ultra-conformable polyimide-based µECoG arrays (referred to as MuSA and CaLEAF going forward) were validated with the best recording configuration to test whether all the electrodes could record the high-frequency components of the evoked responses, independent from their geometry. Finally, two µECoG arrays were acutely implanted in a human patient undergoing awake neurosurgery for tumor resection (low-grade glioma) to investigate speech production processes in speech-related cortical regions. Thus, the ability to recognize speech-related motor intention directly from the neural signals was explored. The neural signals were pre-processed to extract the power spectral features, which were used to train a support-vector machine binary classifier. The performances were then evaluated with standard and non-standard metrics for different combinations of hyperparameters (frequency bands and number of features).

Results. The preliminary results obtained from the setup validation demonstrate that not all the reference-ground configurations tested are indicated for µECoG recording; indeed, the onsite reference might affect the signal quality when it is located too close to the source of interest. This result suggests that the *Single-Ended Screw* is the most appropriate for µECoG recording. This configuration was used also during MuSA and CaLEAF devices validations. Evidence shows that both µECoG arrays can record the high-frequency components of the SEPs, independent from the electrode size, though small electrodes show higher background noise than large electrodes. Finally, for the first time, µECoG grids were acutely implanted in a human patient to investigate speech production processes in the speech arrest area. Recorded neural signals were characterized by

different and well-defined time-frequency components, time-locked to speech production. More specifically, beta, gamma and high-gamma oscillations were involved during speech production at different timings and spatial localization. However, the best performances in the prediction of speech preparation were reported for the high-gamma features.

Conclusion. The results of this work provide new insights into understanding complex neural processes behind speech production that are still not well understood, with a spatial resolution not previously attained in cortical recordings. The µECoG data provide valuable information at a very high spatiotemporal resolution, which could have important implications for the design of speech brain-computer interfaces. In the future, our data from recordings in humans and the related analyses should be confirmed in more subjects to confirm the robustness and reliability of our speech prediction system.



Keywords. micro-ECoG, Spatial Resolution, Broca, Speech Production, High-Gamma

CONTENTS

1. INTRODUCTION

1.1 Principles of Neuroscience

Neuroscience approaches the nervous system from several perspectives of view and at different temporal and spatial scales. Nowadays, the main experimental approaches employed in Neuroscience research can be divided in three principal streams: *in silico, in vitro* and *in vivo. In silico* experiments consist in building detailed multi-scale reconstructions or simulation of brain circuits and their components in order to understand functions and connections of the nervous system. The *in vitro* approach allows investigating specific phenomena or diseases under controlled conditions, mainly at the cellular level. Finally, *in vivo* refers to experiments conducted in animal models or human subjects with the aim of understanding the brain physiology and to eventually test therapies and clinical procedures. These experiments allow also to observe and characterize brain structures, functions and behaviors during specific experimental conditions in either healthy or unhealthy subjects. The common goal of these complementary approaches is to progress our knowledge in Neuroscience, taking inspiration also from advances in other disciplines and applying them to the study of the brain and its functions.

1.1.1 Brain and neurons

The nervous system consists of two main parts defined as the *central nervous system* (CNS) and the *peripheral nervous system* (PNS). Specifically, the CNS mediates the most

basic forms of behavior through two main components: the brain and the spinal cord. The brain controls the activities of the body by processing, integrating, and coordinating the information it receives from the sense organs. It is composed primarily of two broad classes of cells: neurons and glial cells. Neurons are usually considered as the most important cells in the brain [1], not only because of their density (in humans, an overall amount of 10^{11} neurons that is packed in a skull volume of 1.5 litres [2]). However, it is not merely the number of neurons that makes the brain so effective in rapidly processing and reacting to external or internal stimuli. According to literature, the computational efficiency of the brain derives from the connectivity and the communication between neurons and other parts of the body [2]. The unique ability to send signals to specific target cells over long distances makes neurons capable to communicate with other cells located nearby or in distant parts of the brain or body. This fundamental property is due to the presence of a thin fiber, the *axon*, extending from the cell body. Axons occupy most of the space in the brain and projects to other sites with several branches. Neuronal signals are transmitted across the brain in the form of electrochemical pulses called *action potentials*, also known as *spikes*. Spikes propagate from one neuron to another through specialized junctions called *synapses.* The communication starts when a neuron releases the *neurotransmitters*, chemicals stored in the axon synaptic terminals. These messengers bind to receptors present in the postsynaptic neuron at the level of the *dendrites,* ramifications that host several synapses, and the soma. From there, the signal travels along the axon of the second cell to arrive at the axon terminal and propagates across other connections. Thus, together with dendrites, synapses are the key functional elements of the brain. A single axon may make several thousands of synaptic connections with other cells. It has been estimated that the human brain contains approximately 100 trillion synapses. The foundations of sensation, perception, and behavior derive from the information that travels from the periphery *toward* the CNS (or deeper centrally within the spinal cord and the brain) by means of afferent neurons. Inversely, information can be carried *away* from the brain or spinal cord (or away from the hub in question) through efferent neurons. This general distinction divides such networks into two functional systems: the sensory system that acquires and processes information from the environment (e.g., the visual system or the auditory system), and the motor system that

responds to such information by generating movements. The information that travels all over the brain creates neuronal oscillations.

1.1.2 Neuronal oscillations

Neuronal oscillations, or brainwaves, are repetitive patterns of neural activity generated by the neurons or networks of neurons in the brain. Even though the spikes of single neurons in the brain might be difficult to predict, since their activity strictly depends on the connections, their spatiotemporal summation creates rhythmic waves, the oscillations [2]. Techniques such as Electrocorticography (ECoG) and Electroencephalography (EEG) allow to observe and measure large-scale patterns generated by more connected neurons that fire together. The classification of the main frequency bands used in both clinics and research are delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (15-30 Hz), low gamma (30-70 Hz) and high gamma (70-150 Hz). Brainwaves can come simultaneously at different frequencies and they can appear in different regions. For instance, faster rhythms (gamma activity) linked to several cognitive processes, such as speech, can be present in the motor and auditory areas, as well as in more circumscribed cortical regions, such as Broca's area [3-5], the frontal region for speech production. The spatial extent of cortical involvement depends primarily on the frequency of oscillations. As a rule, slower frequencies involve more extensive synchronous activation of the neuronal pool, while high-frequency oscillations usually emerge in more localized regions [2].

In this thesis, I inspected the brainwaves generated by two different models using two different areas under specific experimental conditions. The first network is a well assessed example of cortical functional organization in rats, known as *Barrel Cortex*, studied eliciting *Somatosensory Evoked Potential (SEP)*. The second is *Broca's area*, one of the regions responsible for speech production in humans. These two approaches are briefly described in the following paragraphs.

1.1.2.1 Somatosensory Evoked Potential

The whiskers of mice and rats serve as a highly sensitive tool to detect and acquire tactile information. In fact, rodents build spatial representations of the external environment, locate objects and discriminate different textures by their whiskers [6]. As already

mentioned, the somatosensory whisker-related area, also known as *Barrel Cortex,* is an extraordinary example of cortical functional organization of the brain. The representation of the whiskers follows their spatial localization on the body (see Figure 1 [6, 7]) in a highly organized somatotopic map [6, 8]. Each whisker is represented by a column (or barrel) containing neurons that respond specifically to the stimulation of that specific whisker. The barrel columns can be subdivided into six cortical layers from the more superficial (L1) to the deepest (L6) [9]. All layers include axons, dendrites and synapses. As shown in Figure 1B, the cortical barrels in layer IV of the primary somatosensory cortex [6] are arranged very similarly to the layout of the whisker follicles of the rodent.

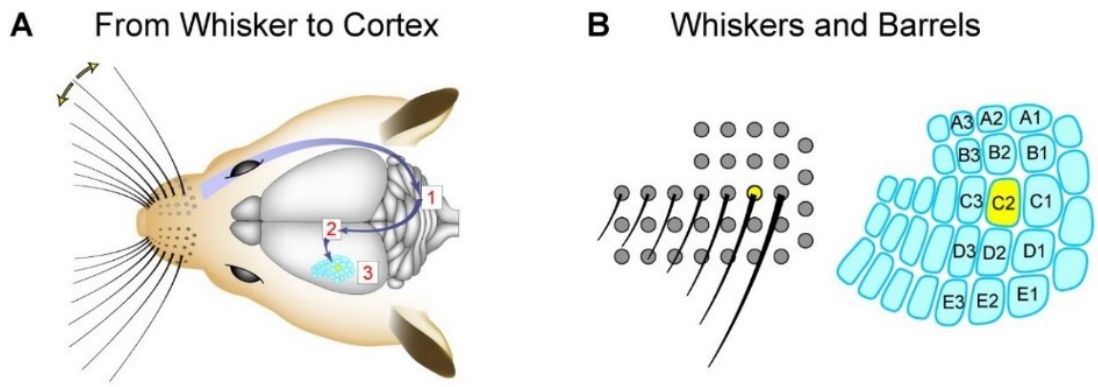


Figure 1. Schematics of the rat whisker-to-barrel system. A) The deflection of a whisker evokes action potentials in sensory neurons. The central whisker C2 is marked in yellow. B) Representation of the cortical columns for the whiskers in layer IV of the primary somatosensory cortex. Each whisker is represented in neurons responding preferentially to its stimulation.

This somatosensory system has been widely investigated in previous studies [6, 8, 10-13], aiming to understand the processing of sensory information in well-defined synaptic pathways. In light of the well-known activity evoked by the whisker deflection (see Figure 2) [11], this model represents a relatively simple but also attractive network to investigate and test new devices [14]. The single-whisker sensory response propagates from neurons in the cortical layer 4 to neurons of layer 2-3 belonging to the same barrel. However, the fast oscillations spread beyond the barrel borders and begin to interact within the barrel field, propagating across the barrel map over the next milliseconds [6, 11]. The primary

somatosensory (S1) barrel cortex connects with the secondary somatosensory cortex (S2) and the primary motor cortex (M1) of the same hemisphere, although callosal projections are also present but less prominently. The stimulation of a single whisker (vibrissae) results in a positive/negative activation in the contralateral hemisphere of the brain (see Figure 2 [15]).

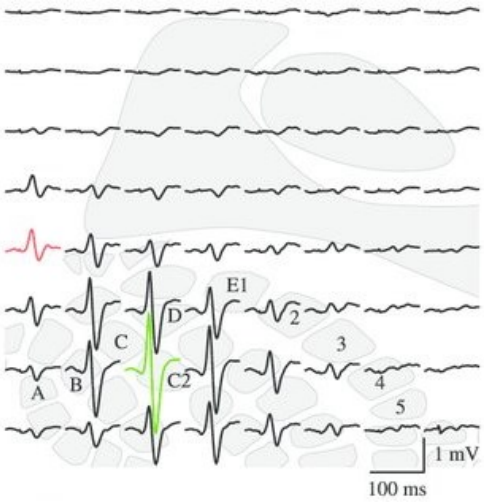


Figure 2. Example of somatosensory-evoked potentials (SEP) averaged across animals, obtained from the cortex in response to stimulation of a single whisker (C2).

1.1.2.2 High-gamma oscillations in Broca's area during speech production

Language-related processes have been deeply investigated and results show that such task recruits several cortical areas, which are widely distributed all over the brain [16, 17]. Various studies provide evidence that high-gamma activity, important during cortical processing [18], is robustly engaged during speech perception and production [17, 19-22]. It is widely recognized that important regions for speech processing are distributed in the parietal, temporal and frontal cortices (see Figure 3 extracted from a review of 2011 authored by Friederici [23]).

Because of its unclear function, one of the most debated language-related regions of the brain is the Broca's area. This area was firstly described by the French neurologist Paul Broca, who observed speech impairment in a patient with a lesion in the posterior part of the third frontal convolution of the left hemisphere, later called Broca's region [24]. This

area, which corresponds to Brodmann's areas (BA) 44 and 45 of the left hemisphere, has classically been associated with language, but its connectivity with other brain regions, as well as its specific functions, still remain unsolved [23].

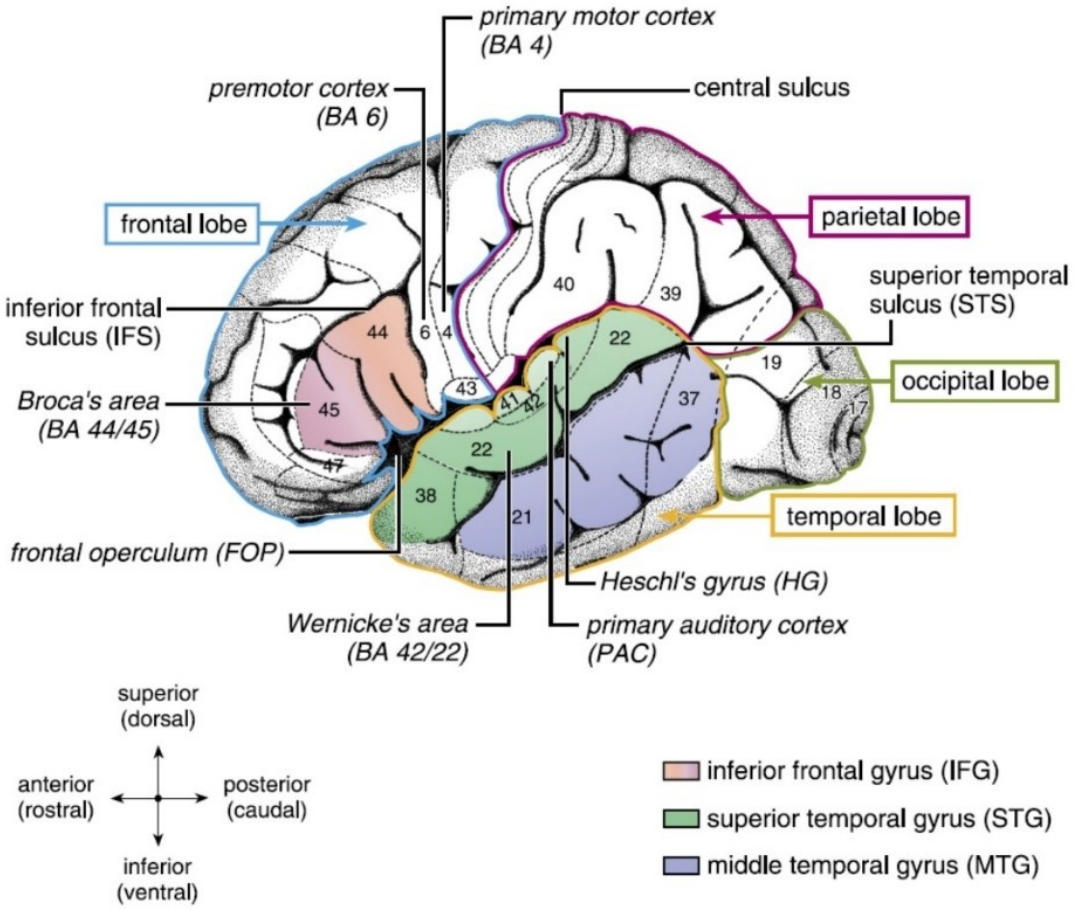


Figure 3. Simplified view of primary motor cortex, premotor cortex, IFG, STG, MTG, involved during speech production and comprehension.

Nowadays, the gold standard for the identification of eloquent areas is the intraoperative functional mapping carried out through Direct Electrical Stimulation (DES) [25-27]. When performed during speech production, this procedure may induce the so called *speech arrest* phenomenon, i.e. the complete interruption of ongoing speech in absence of oro-facial movements and vocal output [28]. This reversible functional arrest allows the identification of Broca's area, and it provides evidence of its critical importance for speech production.

In a recent study conducted by Flinker et al. [5], the role of Broca's area, as well as of the motor cortex and Superior Temporal Gyrus (STG), has been studied during word production with standard ECoG. Results show that electrodes covering Broca's area exhibit an activation in the high-gamma frequency range immediately before the speech onset. Moreover, after speech starts, activations are present in the motor cortex but not in Broca's area.

Another recent study based on DES provided further evidence that Broca's area is involved in early phases of speech production [28]. Results show that no effects are induced by DES during speech production. On the contrary, speech never starts if DES is applied before speech onset. This study provides evidence that Broca's area plays a fundamental role before the onset of speech and thus that it may be involved in pre-articulatory functions.

1.2 State of the Art about speech Brain-Computer Interfaces

As more questions about the functions of the brain arise, the need for more advanced experimental tools arises as well [29]. One recent approach is that of realizing *brain-machine interfaces (BMI)*, or *brain-computer interfaces (BCI),* wherein the brain and an artificial device communicate via electrical activity using an interface. The current brain-recording approaches can show different advantages and drawbacks depending on the purpose. In the next section, I will to discuss the current recording methods focusing on the most suitable for BCI development and speech neuroprosthetics.

1.2.1 Recording brain activity

As already mentioned, one of the essential functions of the brain is communication. The mechanisms that encode neurons firing into perceptions, thoughts and actions have been studied for many decades with several brain recording techniques, each one with its own purpose, level of invasiveness and spatiotemporal resolution (see Figure 4 [30]). These two features become even more important when dealing with the choice of the most suitable method for BCI systems. Indeed, the ideal recording technique for speech neuroprosthetic should be non-invasive and, considering the fast time varying characteristics of the process, able to ensure highly resolute signals in both time and space

[31]. On the one hand, metabolic-based imaging, such as functional Magnetic Resonance Imaging (fMRI), guarantees relatively high spatial resolution and absence of invasiveness, but is affected by low temporal resolution. This technique does not allow to properly record processes that change rapidly. On the other hand, electrophysiological techniques are characterized by higher spatiotemporal resolution, but they still have few disadvantages. Electroencephalography (EEG) is not invasive and easy to set up, but this recording method has low spatial resolution and is highly sensitive to both volume conduction noise and motion artefacts. Magnetoencephalography (MEG) has higher spatial resolution, but it is largely affected by motion artefacts and external magnetic signals. Moreover, it requires a big apparatus, not applicable for practical BCI. These drawbacks make EEG and MEG ill-suited for BCI applications, especially for speech neuroprosthetics. On the contrary, intracortical devices such as microelectrode arrays (MEAs) can record activity with an unparalleled spatial and temporal resolution (single neurons activity) [30, 32, 33]. Unfortunately, MEAs are also the most invasive devices. Since penetration of the brain is required, tissue inflammatory responses and neuronal cells death are elicited affecting in turn the long term stability of the signal [33, 34]. In addition, even though MEAs may allow BCIs for motor prosthetic control [35-37], their spatial coverage is eventually too limited for the decoding of speech, in which the networks recruited are distributed in several brain regions [31]. Less invasive recording approaches such as ECoG (ECoG) can overcome the aforementioned limitations. The ECoG grids of electrodes offer indeed the advantage of recording neural activity directly from the cortical surface with a spatiotemporal resolution inaccessible to non-invasive brain recording techniques [16, 38-40] and reduced invasiveness when compared to intracortical devices [41]. In addition, ECoG recordings are not attenuated and filtered by the scalp and skin and may provide high-frequency neuronal signals [4, 16, 42].

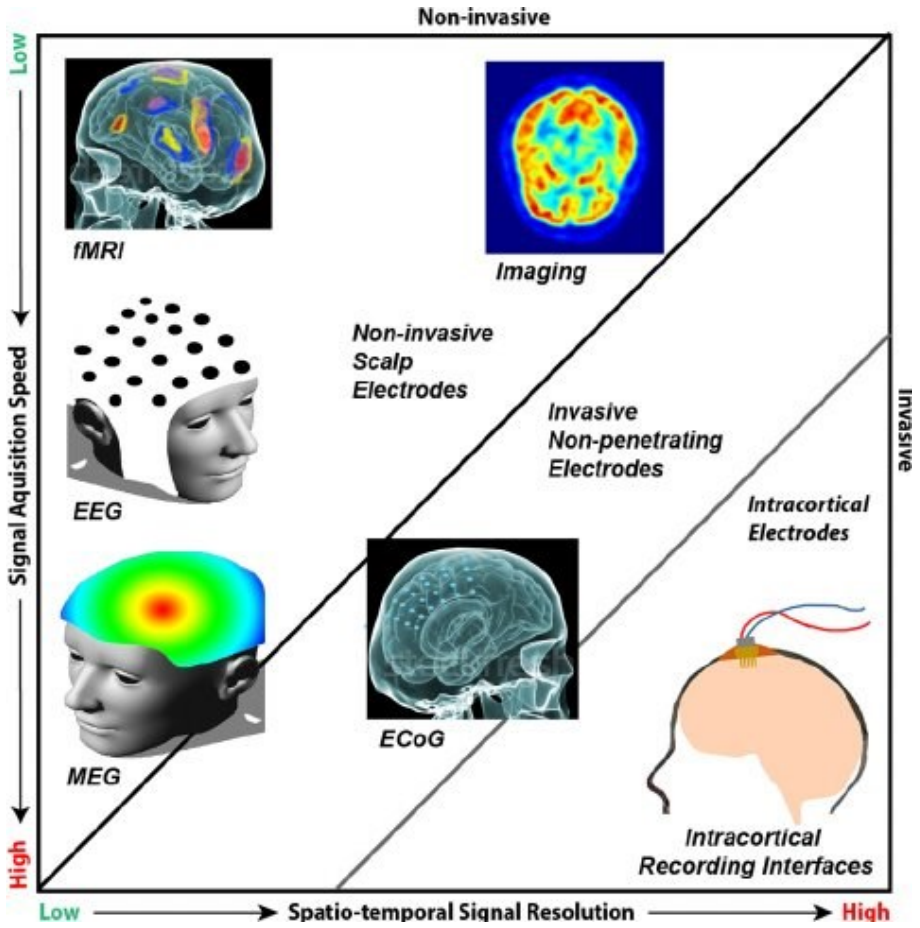


Figure 4. Overview of the current methods used to record brain activity. Brain recording techniques grouped by level of invasiveness and spatiotemporal resolution, from the low invasive with low resolution, Electroencephalography (EEG) and fMRI, through the mildly invasive with high resolution, Electrocorticography (ECoG), and finally to extremely invasive and stereo-encephalography (SEEG) and intracortical single neuron recording (single cell resolution).

The state of the art regarding ECoG devices application into speech BCI development is described in the next subsection.

1.2.2 ECoG-based interfaces for communication

Standard ECoG grids are typically composed of platinum electrodes embedded in soft silastic (silicon) film. The number of contacts goes from few electrodes up to 256 spaced by 1 cm (see Figure 5); individual electrodes are typically 5 mm in diameter, even though they can be of various dimensions. The grids are transparent, flexible, and numbered at

each electrode contact. ECoG grids are surgically implanted either directly beneath the dura (subdural) or above it (epidural) [31, 43]. The placement is guided by the results of preoperative magnetic resonance imaging (MRI).

ECoG grids are considered the gold standard for assessing neuronal activity in patients with epilepsy (where it can be implanted semi-chronically 2 weeks [44]) and are widely used, acutely, for presurgical planning, in order to guide the resection of brain lesions [45-47].

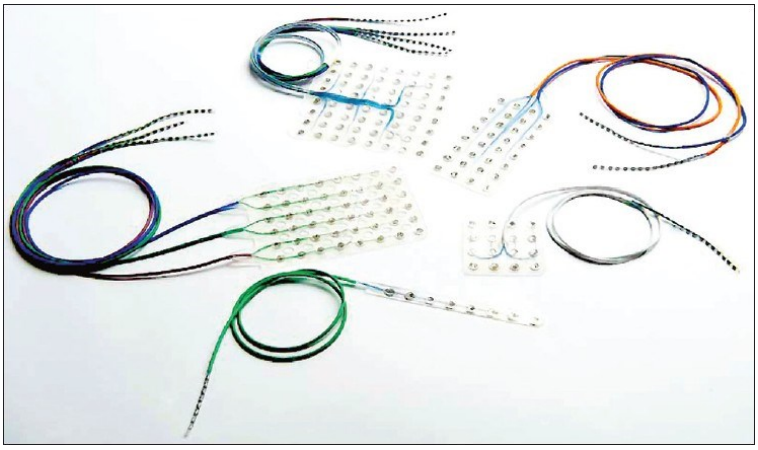


Figure 5. Standard ECoG grid in different shapes and sizes, depending on the clinical applications.

However, this technique has recently emerged as a promising approach for BCI applications to create a direct neural interface that may allow the control of prosthetic, electronic, or digital devices [31, 36, 40, 44, 48-65]. In general terms, all the BCI systems require the following components: a sensor to record and monitor the brain activity and a decoder that converts the brain activity in commands to control the external effector (e.g., a robotic limb, a digital cursor or a virtual keyboard) [66]. Researchers are currently taking advantage of ECoG for speech BCI development [54-56, 67, 68]. A speech BCI can produce different forms of speech output (e.g., words, sentences, synthesized speech) from a pre-processed measure of the user's brain activity. Such technologies aim to restore communication via neural correlates of attempted or imagined speech, improving the quality of life of locked-in patients and others paralyzed patients [31]. Current approaches include also interfaces based on non-invasive recording modalities such as EEG [69].

However, the current gold standards for BCI is ECoG, which has shown much success in the literature in spite of its drawbacks [53, 55, 56, 64, 70].

The neural representation of speech and language is widely distributed, involving several frontal, parietal, and temporal cortical regions. ECoG can monitor different brain areas with excellent temporal resolution, and the population activity captured by this device could provide special advantages for speech BCI [31]. Recently, the possibility of decoding speech from ECoG signals has been deeply investigated using different neural decoders and feature extraction approaches [50, 51, 54-56, 68, 70-73]. Several studies examined the cortical activity related to speech production, which arises in several brain regions at different timings, by means of auditory [39, 52, 64, 67, 68, 74-76], semantic [70, 73] and articulatory features [55, 72]. Very recently, two groups provided evidence that speech can be synthesized mapping the cortical activity into acoustic output, by modelling intermediate representations of production [55, 56]. In particular, Anumanchipalli et al. used a two-stage decoder mapping cortical recordings to synthesized audible speech [55]. The movements of the lips, tongue and jaw, as well as the manner of articulation, were modelled from audio recordings. Starting from vocal tract kinematic model, the first stage decodes the articulatory features from ECoG recording. The second stage maps the articulatory kinematic into acoustic features, such as Mel-Frequency Cepstral Coefficients (MFCCs), voicing and pitch, to reconstruct the participant's speech. To demonstrate that the decoder does not rely only on the auditory feedback of vocalization, and to simulate a setting closer to locked-in syndrome (covert speech), the decoder was also tested on silently mimed speech. Decoded sentences were significantly better than chance-level decoding for both speaking conditions. This study profoundly advances the clinical viability of using speech BCIs to restore communication.

1.3 Main approach

The potential for decoding speech from ECoG signals has been investigated with promising strategies and results [55, 56]. However, several aspects of direct speech synthesis from brain activity need to be optimized before researchers can afford the challenge of clinical trials [31]. A first focus of this thesis was to investigate the possibility of recording high-

frequency components of neural signals from the cortex surface using ultra-flexible micro-ECoG devices (electrode distance below 1 mm, µECoG). Additionally, a second focus of this work was examining implementation of this approach to improve speech-BCI performance in real-life applications.

1.3.1 Micro-electrocorticographic approach to cortical recordings

The first investigation in this work focused on issues concerning tolerability for chronic ECoG implants. It is well known that standard ECoG recording techniques still require invasive procedures due to the dimensions of the grid [31]. Although covering different brain areas recruited in speech processing is an advantage for speech decoding, this traditional approach could be improved by using multiple high-density micro-grids directly placed over functional brain regions of interest [77, 78], instead of only one macro-grid covering several areas of the brain. This innovative technology would increase the spatial resolution of the recording, maximize the specificity of the recorded signals and, at the same time, minimize the risk of damage to the cortex. Moreover, the long-term performance of invasive BCI systems, which strictly depend on the inflammatory response and thus on the size and stiffness of the device [29], would increase substantially. This has been confirmed by empirical evidence showing that probes with a degree of flexibility comparable to brain tissue can minimize foreign body reactions and improve implant biocompatibility [29, 79, 80]. In the present work, I will first describe the results of cortical recordings performed in rats with different micro-ECoG devices, with the goal of validating this approach. Subsequently, I will report the results from micro-ECoG grids acutely implanted in a human patient undergoing tumor neurosurgery to investigate speech-related processes recorded from the left frontal region of the brain.

1.3.2 Micro-electrocorticography during speech production

The second focus of my thesis extends beyond the limits of technical implantation. Speech neuroprosthetic devices aim to improve the quality of life of people suffering from communication deficits due to severe motor limitations. In such patients, even though communication is no longer possible due to severe paralysis of motor cortices [81], language-related cortices might still be intact. Ideally, a speech-BCI system should be able

to decode speech directly from brain activity elicited in eloquent areas for speech production and perception [50, 55, 56, 64, 65, 67, 68]. Detecting speech-related motor preparation would thus be a step toward building a speech decoding system. Indeed, since vocal cues may currently be employed to activate the most common virtual assistants (e.g., Google Assistant, Alexa, Siri), a neuronal cue could serve to initiate speech decoding. Here, I investigate the possibility of recognizing speech-related motor preparation from µECoG data recorded from a patient performing a naming task to provide new insights into understanding the complex and still unclear neural processes of speech production at a very high spatial resolution [82, 83]. I will demonstrate that µECoG signals recorded from speech-related brain regions can provide the necessary information to detect speech initiation. This achievement could be employed as the trigger when a measurable speech event, and thus an explicit articulatory alignment, is not possible (e.g., in the case of covert speech) to start the decoding in real-time BCIs.

2. MATERIALS AND METHODS

In this chapter, I will describe the experimental setup used to record electrocorticographic data either in rats or human subjects (see section 2.1). Afterwards, I will describe the *in vivo* experiments conducted for setup and electrode validation in rats and humans (see sections 2.2.1 and 2.2.2), as well as the specific pipelines followed during the data analysis (see sections 2.2.1.2 and 2.2.2.3).

2.1 The recording setup

The custom-made µECoG devices were connected to a TDT system (Tucker-Davis Technologies), described in the next subsections.

2.1.1 Tucker-Davis Technologies (TDT) based setup

µECoG data were recorded with a multi-channel amplifier optically connected to a digital processor from Tucker-Davis Technologies. Specifically, the main components of the recording setup are the following:

- Headstage with µECoG device and ZIF-CLIP;
- Preamplifier (PZ2);
- Digital signal processor (RZ2);
- Computer;

- Microphone or stimulator.

These components are connected as shown in Figure 6. The headstage is connected via ZIF-CLIP to a battery-based preamplifier, the PZ2. The preamplifier interconnects the device to the RZ base station through a fast fiber optic connection. The PZ2 channels are organized into groups of 16 channel banks, and the signals are amplified, digitized and transmitted to the RZ2 base station. Therefore, the fiber optic connection leaves the experimental subject electronically isolated for safety reasons. The RZ2 processes signals from additional devices (i.e., the microphone) and external triggers. A trigger is an externally generated signal to capture specific events or conditions during a task, such as the timing of a stimulation or different phases of an experimental protocol (listening or speech production). All the data recorded and processed by the RZ2 are stored in the computer connected to it.

2.1.2 The µECoG devices

The data analyzed in this thesis were recorded with three different devices. The Epi device was designed at the Italian Institute of Technology (IIT) and was tested and described in previous studies [14, 84, 85]. Flexible Printed Circuit Technology reduced the size of recording sites and inter-electrodes spacing while using nanostructured gold coatings to ensure low electrodes impedances. The device consisted of 64 recording sites of 140 µm diameter arranged in an 8 × 8 grid with 0.6 mm spacing, producing a total covered area of 4.3 by 4.3 mm. An image of the device with its dimensions is shown in Figure 6A**.** The second and the third devices, the MuSA and the CaLEAF, are ultra-conformable polyimide-based µECoG arrays (see Figure 6B-C). The MuSA devices were produced at the Institute of Microsystem Technology (IMTEK), Laboratory for Biomedical Microtechnology, in Freiburg, Germany, by Maria Vomero and Maria Francisca Porto Cruz. The CaLEAF devices were produced at the same institute by Maria Vomero. The fabrication process for the devices has been previously described [86, 87]. The main differences between the two devices consist of the material, number and diameter of the electrodes. The MuSA device holds 32 platinum (Pt) electrodes, 16 large size (diameter of 100 µm) and 16 small size (diameter of 10 µm); the large and the small contacts are separated by 45 µm. The CaLEAF device consists of 16 glassy carbon electrodes with following sizes: 50, 100, 200, and 300 µm

(four electrodes for each size). The MuSA and the Epi were tested in both humans and animals, while the CaLEAF was only tested in rats.

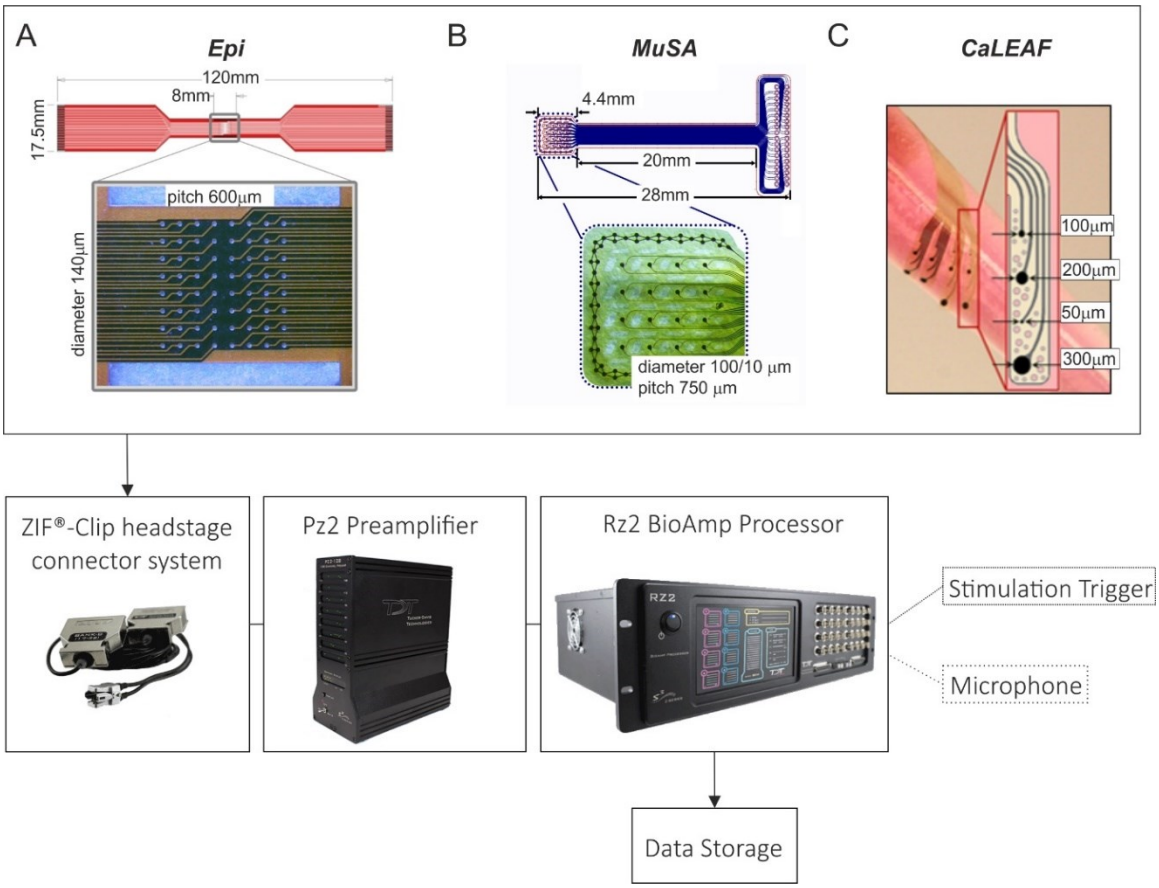


Figure 6. Overview of the setup used during ECoG recordings in both animal and human in vivo experiments. The Epi (A), MuSA (B) and CaLEAF (C) devices have been tested following the chain from the headstage to the processor and computer. The setup can record also external triggers to align the trials during the offline data processing. For in vivo animal experiments, the stimulation trigger was recorded to align the trials to the timing of the mechanical stimulation. During the experiments in human, the external input used to align and analyze the data was the voice of the patient, recorded with a microphone at 24kHz.

2.1.3 The configurations

Recording with a TDT-based system allows the user to choose between three possible configurations in terms of reference (if any) and ground: *Single-Ended Screw (*or *Dura* in human subjects*), Differential* or *Single-Ended Reference* (see Figure 7). During the first part

of the study, I focused my attention on determining the configuration to optimize the recording quality and the evoked-to-spontaneous ratio. Ground and reference placements are important in all headstage configurations because they can produce undesired results if incorrectly wired. The *Single-Ended* configuration requires a connection between reference and ground at the headstage level (see Figure 7B or C). For the *Single-Ended Screw* configuration (C), reference and ground are attached to a skull screw (or clipped to the dura mater), while the two pins are attached to one or more onsite reference electrodes in the *Single-Ended Reference* configuration (B). In the *Differential* configuration (A), the ground is not connected to the reference at the headstage level; in this case, the reference is connected to the onsite reference electrodes, while the ground is attached to the skull screw. Specifically, in the Epi and the MuSA devices, the onsite reference is close to the recording electrodes. The Epi is provided with two lateral reference electrodes, while the MuSA's reference consists of a crown of interconnected electrodes placed around the active area (see Figure 6A-B). The choice and test of the reference electrodes are important because any activity detected by the reference is reflected as activity in all the recording electrodes—that is, the voltage values recorded from each electrode are relative to a voltage recorded elsewhere (reference site).

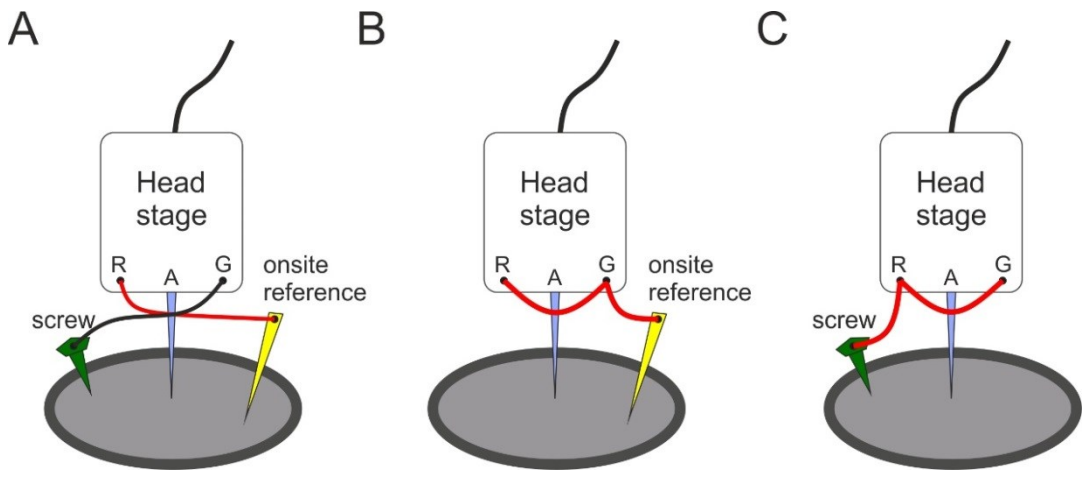


Figure 7. Ground and reference configurations for the headstage. A) Differential configuration: reference and ground are separated at the headstage level, and the ground is shorted to the skull screw (or dura), while the reference is connected to a different site on the subject. B) Single-Ended Reference configuration: ground and reference pins are connected at the headstage level and

shorted to the onsite reference electrode. C) Single-Ended Screw configuration: ground and reference pins are connected at the headstage level and shorted to a skull screw.

2.2 Setup validation: from animal models to humans

2.2.1 In vivo validation in animal models

The experimental procedures to record the data in rats were planned and conducted by Elena Zucchini, while I conceptualized and developed the software necessary for the data pre-processing, analysis and visualization of the results.

2.2.1.1 Recording setup and experiments

First, the team validated the three recording configurations with both the MuSA and the Epi devices, with four and two *in vivo* animal experiments respectively. Next, five CaLEAF and eight MuSA devices were tested for high-frequency recordings in rats.

Neural signals were recorded using the Tucker Davis Technologies (TDT) multi-channel recording system described above; data were digitized at a sample rate of 12207 samples/s at 18-bit resolution. In order to reduce electromagnetic noise, the experimental setup was placed in a Faraday cage.

The experiments for setup and electrode validation were conducted using adult Long Evans rats (400–500 g). To record somatosensory evoked potentials, animal surgery and implantation of the ECoG electrodes over the rat barrel cortex were carried out (A-P: -1 mm to -4 mm from bregma, M-L: -3 mm to -5 mm from the midline) following the procedure described in a previous study [88]. The experimental plan was designed in compliance with the guidelines established by the European Communities Council (Directive 2010/63/EU, Italian Legislative Decree n. 26, 4/3/2014), and the protocol was approved by the Ethics Committee for animal research of the University of Ferrara and by the Italian Ministry of Health (authorization n 332/2015-PR) [87].

For each rat, several positions and orientations of the ECoG devices over the barrel cortex were investigated. The somatosensory evoked potentials (SEPs) were mechanically evoked

with a vibrating system used to produce a whiskers deflection along the horizontal plane. The stimulation involved either several whiskers or a single whisker. The whiskers contralateral to the craniotomy were shortened and inserted into a velcro strip attached to a rod. A shaker (Type 4810 mini shaker; Bruel & Kjaer, Naerum, Denmark) controlled by a National Instruments board (Austin, TX, USA) moved the rod to deflect the whiskers. For single-whisker stimulation, the whisker of interest was inserted into a needle attached to the rod. In both single and multi-whiskers stimulation protocols, the deflection stimulus, consisting of a sine waveform of 12 ms duration, was delivered at 10 Hz. The stimulation amplitude was coincident with a multi-whisker deflection of 500 µm. Each deflection stimulus was repeated 100 times and was separated from the others by a four-second pause.

2.2.1.2 Task-specific pipeline for Somatosensory Evoked Potentials signal processing

Band-pass filtering and segmentation

The processing began with organizing the raw data and experiment information. Each electrode was visually and qualitatively inspected, searching for artefacts or external noise. Raw data were band-pass filtered into different frequency bands using a digital zero-phase 8th-order Butterworth filter to extract different components of the SEP. Data were band-pass filtered between 70 and 300 Hz for the setup validation and between 200 and 1000 Hz for the device validation. The filtered data were segmented into trials of 80 ms, which were time-locked to the stimulation (ranging from 20 ms before the onset to 60 ms after it). The filtered trials were then processed in several steps and inspected in order to understand the spatial distribution of the evoked potentials.

Time-frequency analysis

For the bands of interest, the time-frequency analysis was performed by applying the built-in *spectrogram()* Matlab function, based on the short-time Fourier transform (see A.1). The time-frequency analysis is applied by setting two parameters: the window length and the overlap. The single-trial spectrograms were computed by setting a window of 10 ms and an overlap of 9 ms. For the frequency bands 70-300 Hz and 200-1000 Hz, the frequency step was equal to 5 and 10 Hz, respectively. The single-trial spectrograms were

inspected to exclude the presence of artefacts and were subsequently averaged to obtain mean spectrograms.

Evoked-to-Spontaneous Ratio and statistics

The amplitudes of SEPs evoked by peripheral stimulation of the whiskers were used as measures of the recording quality by calculating the *Evoked-to-Spontaneous* Ratio (ESR). The ESR is usually defined as the ratio of the amplitude of a response evoked by a stimulus to an average value of spontaneous activity and was computed as follows:

$$ESR = median(\frac{PeakToPeak}{2 \cdot std(noise)})$$

where *PeakToPeak* is the single-trial peak-to-peak amplitude of the signal of interest (the evoked potential), i.e., the difference between the maximum positive peak and the maximum negative peak. The *std(noise)* is the standard deviation of the spontaneous activity (100 ms duration before each stimulation), which was considered to be the background for each trial.

The possible difference in ESR performances between different electrode sizes in the MuSA device were tested using a one-way ANOVA. The purpose of a one-way ANOVA is to determine whether data from different groups (or levels) of a factor have a common mean [89]. This analysis enables us to determine whether the groups of an independent variable (here, the electrode size) have different effects on the response variable. This test was conducted only for the MuSA ESR values. Comparison of single-trial ESR values was performed for each large electrode and its corresponding small one (45 µm far); thus, for each recording session, the ANOVA was performed sixteen times (for the sixteen electrode pairs). The test was not conducted for the CaLEAF because a direct comparison across different electrode sizes is not possible; the distance between electrodes is 400 µm, thus the peak-to-peak amplitudes might change with the distance of the electrodes from the signal source.

Interpolated maps from peak-to-peak amplitude

In order to have a better understanding of the spatial distribution of the signals, a bidimensional map was computed by interpolating the average peak-to-peak values

between the electrodes, with a resolution of 30 µm. The interpolation was performed applying the thin-plate smoothing spline method, which is suited to extrapolate surfaces starting from a grid of values. The advantage of this method is that it is also possible to interpolate values outside the starting grid and not only inside. The median evoked potential was also shown overlaid on the bidimensional map.

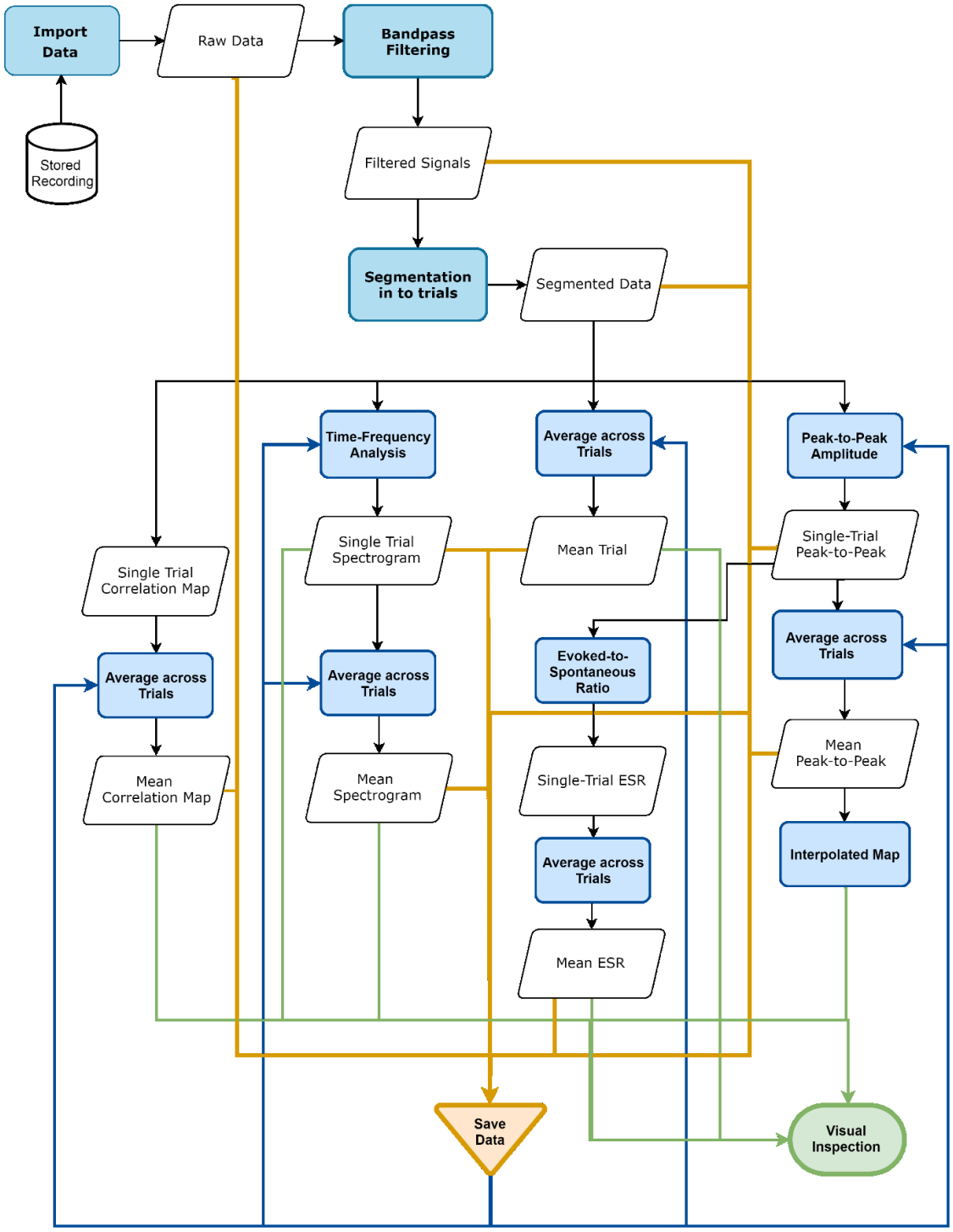


Figure 8. Signal processing pipeline for the SEPs analysis. The black arrows show the direct process flow. The black parallelograms represent the data resulting from the analytical steps (rounded rectangles), which are saved for future elaborations (orange arrows). The light blue arrows indicate the possibility to load stored data at any point in the flow. The green arrows show which type of data can be visually inspected using dedicated functions.

2.2.2 In vivo validation in humans

The data recording in humans was conducted by Dr. Tamara Ius. The analyses were conceptualized by Aldo Pastore and me, while I developed the software necessary for the data pre-processing, analysis and visualization of the results.

2.2.2.1 Participants

Data were collected from one male patient: an Italian native speaker, undergoing awake neurosurgery for tumor resection (low-grade glioma). The patient gave his informed consent, and the protocol was approved by the Ethics Committee of Azienda Ospedaliera Universitaria Santa Maria della Misericordia (Udine, Italy) and by the Italian Ministry of Health.

2.2.2.2 Recording setup and task

Device specifications and recording setup were previously described by Rembado et al. and Vomero et al. [85, 86]. Briefly, two devices were used for the recordings. The first array (Epi) consisted of 64 channels arranged in an 8x8 layout, with a pitch of 600 µm and a diameter of 140 µm (see Figure 10A-left); the second array (MuSA) consisted of 32 channels, among which only the 16 with a diameter of 100 µm were considered. The MuSA electrodes are arranged in a 4x4 layout, with a pitch of 750 µm (see Figure 10B-right). Array references were disconnected, and the ground was connected to the *dura mater* of the patient (Single-Ended Dura configuration).

The patient's voice was acquired at 24 kHz for the entire recording sessions, and neural signals were collected before the tumor resection procedure. The placement of the µECoG devices over a healthy and informative portion of the cortex was obtained with preoperative and intraoperative mapping. The patient underwent functional Magnetic Resonance Imaging (fMRI) while performing a naming task (see Table 1) as part of his clinical plan to localize the responsive areas and their distance from the tumor. Subsequently, to identify the speech arrest area and center the µECoG arrays above it, cortical mapping was carried out with intraoperative DES. To determine the position of the devices above the cortex, the four corners of the Epi and one point of the MuSA (due to its small dimensions) were acquired with a neuronavigator. Since only one point was

acquired for the MuSA device, their relative position was estimated by plotting the coordinates in two dimensions (see Figure 9).

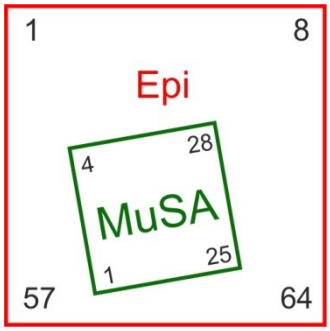


Figure 9. Relative position and orientation of the MuSA and the EPI devices. The numbers reported in the corners correspond to the electrodes in those positions.

The subject was asked to repeat the same naming assignment performed during the fMRI scan. This procedure is routinely used with this task or other similar tasks (i.e., a counting task) to monitor speech functionality during the tumor removal surgery.

The task consisted of naming different images displayed on a screen three times in the same order. In total, 30 trials with the Epi and 30 trials with the MuSA were recorded in separate and consecutive recording sessions. The Italian nouns repeated during the task are reported in Table 1.

Table 1. Naming Task performed during the recordings (first row) and the English translation (second row).

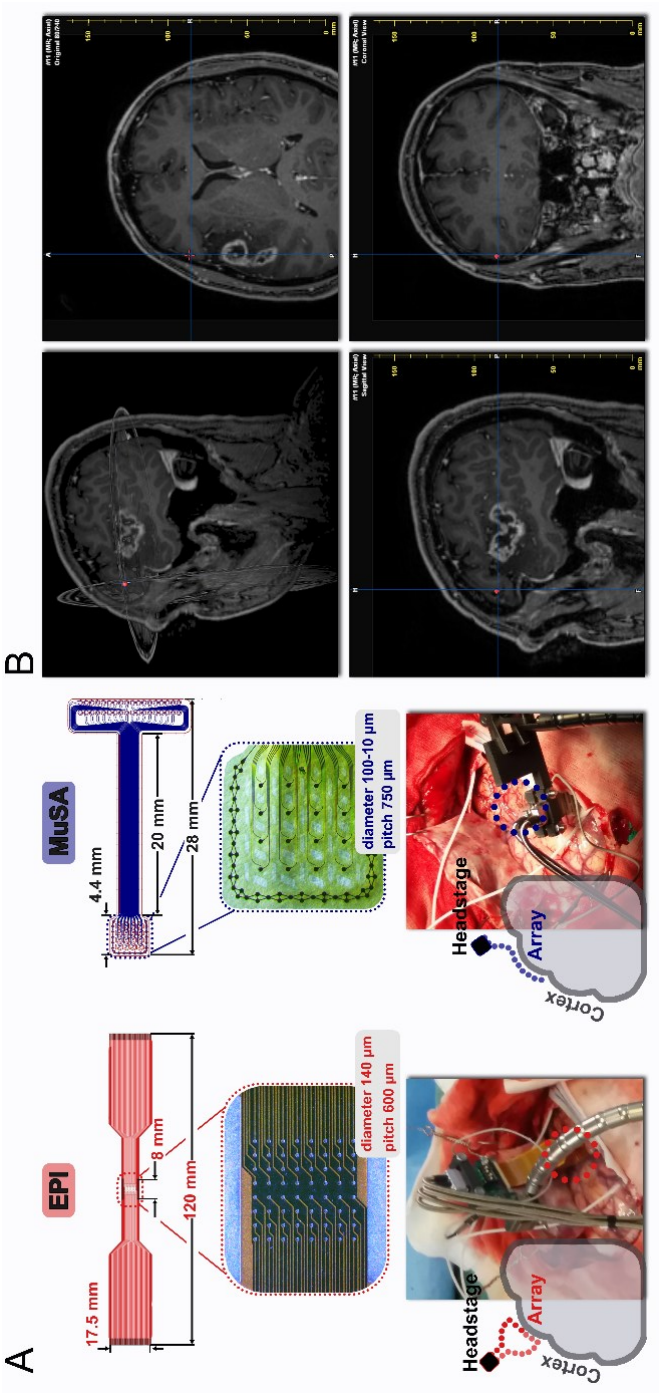


Figure 10. Overview of the devices and their positions over the cortex during in vivo experiment in humans. A) Schematics and a pictures of the active sites area of the two micro-ECoG arrays used for the recordings (A-left Epi: electrodes diameter 140 µm - electrodes pitch 600 µm; A-right MuSA: electrodes diameter 100 µm - electrodes pitch 750 µm). Pictures of the devices over the patient cortex are shown in the bottom. The superimposed sketch represents how differently the two arrays adapt to the cortex. B) Snapshot from the neuronavigation system showing the µECoG array location superimposed to the MRI scan. The four boxes show different brain sections (clockwise starting from the top left image: 3D view, horizontal plane, coronal plane and sagittal plane). Red dot corresponds to the speech arrest area and thus to where the arrays were placed.

2.2.2.3 Task-specific pipeline for predicting speech preparation from µECoG signals

The pipeline followed during the offline processing is reported more in detail in Figure 11. The processing started with loading and organizing the raw data and the experiment information. Raw data were band-pass filtered into the traditional beta (15 - 30 Hz), gamma (30 - 60 Hz), and high gamma (70 - 150 Hz) bands. The Matlab function filtfilt() was applied to avoid phase distortion (50 Hz; harmonics up to 150 were removed with notch filters). The filtered data were subsequently segmented into 1 second trials time-locked to the speech onset (ranging from 500 ms before the onset to 500 ms after it) for visual inspection. Last, the filtered trials were processed and inspected to check the quality of the data. The visual inspection consisted mainly of correlation analysis and time-frequency analysis (see sections A.1 and A.2).

Correlation analysis

Correlation analysis is a normalized measure of covariation that quantifies the strength of a relationship between two variables. It is usually applied to find a linear dependence between measures and to explore network function and connectivity [90].

The correlation of neural signals was analyzed in two ways:

- Electrode correlation as a function of the pitch: this measure is often used both as an index of spatial propagation of neural signals and as a metric to define the resolution of particular phenomena in specific frequency bands [77, 91]. The Pearson correlation coefficient was computed trial by trial from the filtered signals and then averaged across electrodes sharing the same distance; the result is a mean profile for each inspected frequency band.
- All-vs-all electrode correlation: a mean correlation map for each electrode compared to the others was computed starting from the single trial and then averaged across trials for the frequency bands of interest.

The mean correlation profile and the correlation maps were explored to investigate the redundancy of the information as well as to define the spatial distribution of task-specific

dynamics. The correlation maps and profiles were computed for beta, gamma and high-gamma ranges and plotted for visual inspection.

Time-frequency analysis and mean power profile

For the bands of interest, the time-frequency analysis was performed by applying the built-in *spectrogram()* Matlab function based on the short-time Fourier transform (see A.1). The spectrogram computation was performed considering a window of 100 ms and an approximate overlap of 90 ms, while the frequency step was adapted to the range of interest. Next, the mean spectrogram for each electrode was obtained averaging across trials. The single-trial and mean spectrograms were inspected to exclude the presence of artefacts.

The mean power profile (MPP) was calculated starting from the spectrograms, computing the average across frequencies, to obtain a mean time-varying profile of the power spectral density.

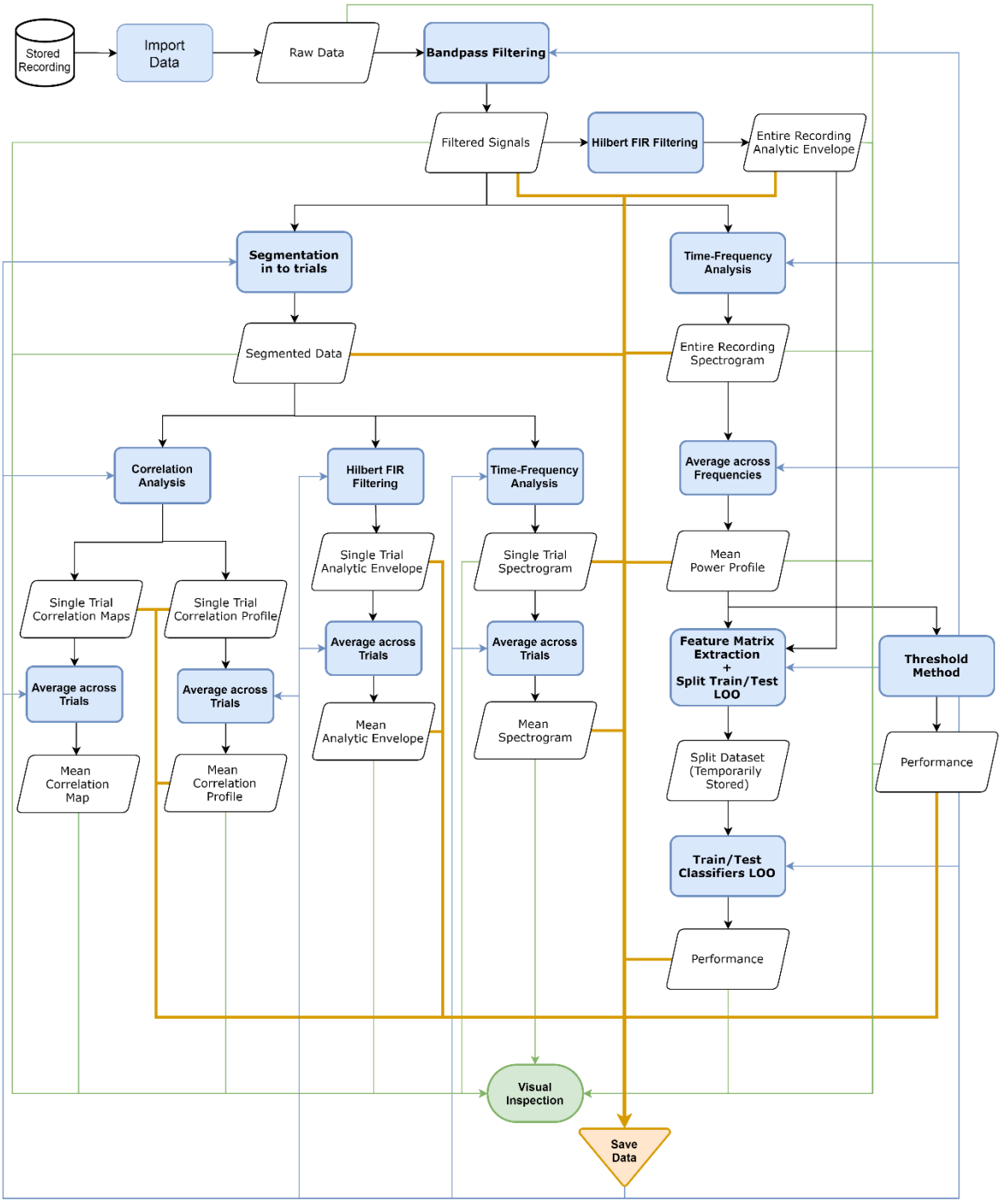


Figure 11. Signal processing pipeline followed for the speech production analysis. The black arrows show the direct process flow. The black parallelograms represent the data resulting from the analytical steps (rounded rectangles), which are saved for future elaborations (orange arrows). The light blue arrows indicate the possibility to load stored data at any point in the flow. The green arrows show which type of data can be visually inspected with dedicated functions.

Feature extraction for speech preparation classification

To obtain a matrix of features, the MPP was computed for the entire recording session and then divided in 30 intervals, each containing only one naming execution. Specifically, segments ranged from 500 ms before the speech onset of the included word to 500 ms before the onset of the subsequent one. The observations provided to the classifier were obtained by considering each electrode independently. This approach uses two hyperparameters (see Figure 12): the length of the window *w*, and the length of the shift *s.* Every time point included in a window was considered to be a feature; thus, its length determines the size of the feature vector. The number of overlapping time points, and thus the number of features that were present both in one observation and in the subsequent one, was established by the length of the shift. The observations are considered to be speech preparation when comprised between 500 ms before the speech onset and the onset (class 0). Instead, when the entire window the observation belong to the speech production segment (after the speech onset) class 1 is assigned (non-preparation class).

Formally, with a shift length s and window length w, each observation obs_j ranges between $[t_j, t_j + w]$, and the subsequent observation starts at $t_j + s$ and ends at $t_j + s + w$. This approach offers the advantage of providing temporal context information to the classifier, a common method described in previous literature [54, 92]. In our experiment, the shifting window covered approximately 60 ms, and no overlap was set ($s = w$). For this set of parameters, the observation matrix consisted of nine observations for the preparation class and a variable number of observations for the non-preparation class, according to the temporal distance between a trial and the consecutive one. In order to train the classifier, each observation of i-th interval was matched with a binary label. If $t_{onset_i} < t_j < t_{onset_{i+1}} - 500ms$, the observation was labelled as 1. When $t_{onset_i} - 500ms < t_j < t_{onset_i}$, the observation belonged to the preparation interval and was labelled as 0.

We assume that the power spectrum of the selected frequency band is characterized by well-defined temporal patterns representing speech preparation that did not emerge at other time points. However, these anticipatory neural patterns might carry some

physiological variability in timing, which could lead to wrong labelling when a precise and fixed temporal window is used to assign the classes. For the sake of simplicity, I followed the procedure indicated above, but this is likely to have made the classification task harder.

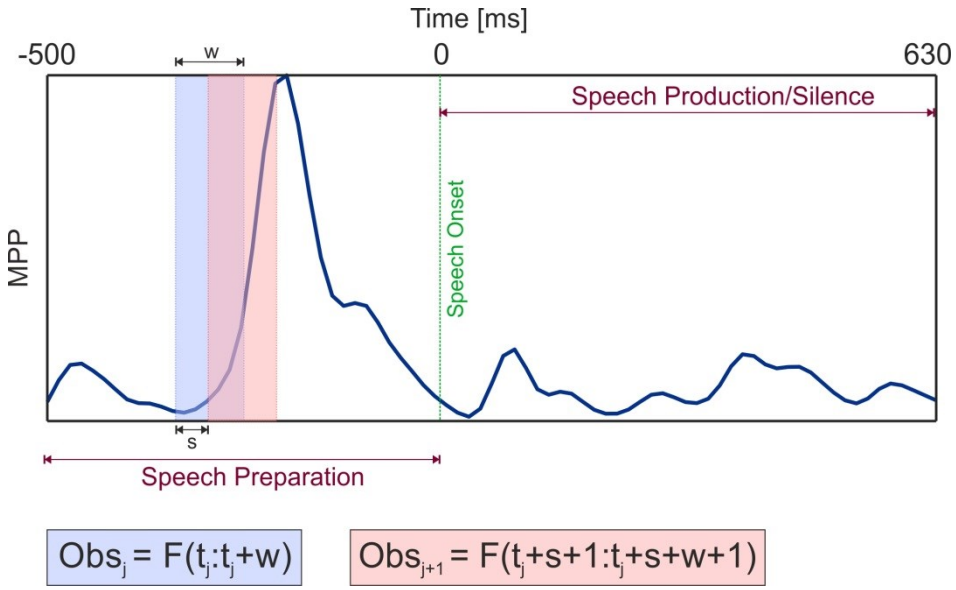


Figure 12. Feature matrix creation and labelling procedure starting from the features F (MPP or Envelope). Considering a shift length s and context length w, and considering an observation obs_j ranging between $[t_j, t_j + w]$, the following observation would range $[t_j + s, t_j + s + w]$. One segment starts 500 ms before the speech onset and ends at 500 ms before the speech onset of the subsequent trial. The observations are considered to be speech preparation when comprised between 500 ms before the speech onset and the onset (class 0). Instead, when the entire window belongs to the speech production segment (after the speech onset) class 1 is assigned.

Dealing with unbalanced classes

Most real-world classification problems are characterized by imbalanced classes—that is, the two classes are not equally represented in the dataset. This data imbalance problem is recognized as a major issue in the field of data mining (see A.4). Indeed, most machine learning approaches assume that data are equally distributed; if not, the majority classes would dominate over minority classes, causing classifiers to be more biased towards majority classes, with poor or zero detection of minority classes. Another consequence is difficult performance evaluation, since overall accuracy is

not a good criterion to assess the classifier's performance in imbalanced domains. The goal of recognizing speech preparation is a clear example of a real-world classification problem suffering from imbalanced classes. Possible approaches to dealing with imbalanced classes consist of re-sizing the data either by *undersampling* or *oversampling* [93] in order to change the prior conditions of the training data. Here, considering the small amount of data, overfitting was likely to occur with oversampling; therefore, I decided to adopt the undersampling approach, a popular technique for unbalanced datasets that reduces the skewness in class distributions.

The undersampling procedure consisted of reducing the number of observations belonging to the speech/silence class by using a *proportionality factor (pf).* By definition, this factor depends on the majority/minority ratio. However, due to the leave-one trial-out approach, the data segmentation could produce trials of different length, as described above. In this case, a definite *pf,* computed as the ratio #majority/#minority, could be at the most equal to two, because of the presence of some very short trials. Such balancing between the two classes would probably improve the classification performances when dealing with large datasets by reducing the prior probability of the majority class. However, in this case, it would also drastically reduce the number of observations. Therefore, I decided to test different values of *pf* consisting of the maximum imbalance allowed. The tested *pf* were integer values ranging between 1 and 5. Finally, also the uncontrolled imbalance was tested. To control the level of imbalance, the majority class was randomly downsampled during the training: briefly, if the minority class had *k* observations, the majority class could have at the most *pf·k* observations, according to the length of the segments. This downsampling allowed for controlling the proportions of the minority and the majority class, as well as finding the best trade-off between undersampling and balancing.

This test was conducted for the Epi recording device, since the task-related power appeared to be more consistent in this device and was present in more electrodes. Subsequently, the best combination was then tested with the MuSA recording device.

SVM classification

To increase the reliability of the results, I used a leave-one trial-out approach. The recording session was segmented into 30 intervals, and all the observations located inside 29 of 30 intervals were used to train a binary SVM classifier with a Gaussian kernel (see A.5). The classifier was then tested using the observation from the excluded segment. The training procedure was repeated separately for each channel and frequency band of interest (beta, gamma and high-gamma) to understand if the decoding was operating better in some combinations of spatial localization and frequency modulation. For completeness, the classifier was also trained with randomized labels for the best combination of window, pf and features.

Classifier performance

Unfortunately, one of the main complications of this study was that the evaluation of our classifier performance was derived from the assignment of an ambiguous class such as speech preparation. However, similar to covert speech, there is little information about the real timing of preparation intervals, which implies inevitable errors during the labelling procedure [68]. This problem is already known in literature as *label noise* [94, 95]. Therefore, the goodness of the classifiers cannot be evaluated only with standard measures (see A.6); other evaluation metrics should also be taken into account. For these reasons, one reliable measure to quantify the prediction accuracy could be provided instead by the number of detected trials, as explained in the following paragraph.

To assess how well a trained model was performing, an index of performance was defined taking into account both the percentage of identified preparation segments and the number of false positives. Each preparation was considered to be detected if the classifier was able to correctly identify at least *n* observations as preparation, for different values of *n* considered (between 1 and 4), inside the corresponding fixed time intervals (from 500 ms before the speech onset to the speech onset). Accordingly, the performance of the classification index (PCI) can be described as follows:

$$PCI = 100 * \frac{\#\text{Preparation}_{\text{detected}}}{\#\text{Preparation}_{\text{total}}} - 1.5 * \frac{\#\text{NotPreparation}_{\text{misclassified}}}{\#\text{NotPreparation}_{\text{total}}}$$

To achieve the highest positive score, the decoder must detect all the preparation intervals without misclassifying the observations of the other class. However, if no preparation is detected and every non-preparation observation is misclassified, the most negative value is obtained. In addition, the *F-score* and the *Matthews correlation coefficient* were also computed and used as evaluation criteria (see A.6).

2.2.3 Graphical User Interface for ECoG signals processing

To identify key array features underlying high-quality signals, I analyzed several datasets recorded with high-density µECoG devices. In order to provide a tool for systematically processing and analyzing the acquired neuronal signals with repeatable methods, I developed a user-friendly graphical interface (GUI). The GUI and the analyses were implemented in Matlab (version 9.5, Mathworks, Inc., Natick, MA), with the only exception of the speech timing computation, which was extracted with the free software Audacity(R) (version 2.1.2). Audacity® software is copyright © 1999-2019 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni. All the analyses described in the previous sections, which were carried out both to inspect and process the data, have been included in the GUI and organized into several menus, which allow the user to control the procedures and the parameters. In particular, the main parameters for trials segmentation (segmentation timings), filtering (frequency range, sampling frequency), spectrogram computation (window and overlap) and feature extraction (window and overlap) are controlled by the user; during the processing, this information is added through popup dialog boxes.

The first step to begin analyzing a new dataset is loading the raw data, including the information about the experiment. The user is asked to import the following information:

1. The identification numbers of the blocks to be analyzed;
2. The number of different devices used during the experiment;
3. The type and the name of the devices;
4. Whether a trigger has been recorded during the experiment or not; if yes, either the on-board analog input number of the trigger or the .txt files where the timings are stored will be requested;

5. The minimum number of samples that should elapse between one trial and the following one.

The information provided will be stored in a file, which will set the starting point of each further analysis in order to reduce the computational cost. For example, after filtering and segmenting the data, the resulting matrix will be stored in a file, which can be loaded in the future to apply visual inspection algorithms as well as exploratory analyses.

The SEPs analyses as well as the speech prediction were conducted entirely through the GUI to guarantee repeatable measures and results for future experiments.

3. RESULTS AND DISCUSSIONS

In this chapter, data analysis of the *in vivo* validation of the setup (see section 3.1) are reported along with the MuSA and CaLEAF validation evidence (see sections 3.2 and 3.3). Next, findings obtained from translational application of the Epi and MuSA devices in one human subject is described (see 3.4). I will provide evidence that neural signals recorded with sub-millimeter-spaced electrodes exhibit low correlation and that different frequency bands show diverse and well-defined temporal dynamics and spatial localizations, strictly depending on the experimental conditions. Furthermore, I will explore whether it is possible to exploit high spatial resolution to understand speech production processes in the speech arrest area. Finally, I will report the findings of the prediction of speech production, obtained using high-gamma activity anticipatory features extracted from the neural signal.

3.1 Setup validation: improving recording quality

3.1.1 Results

The setup was validated with the MuSA and the Epi devices in several experiments. For simplicity, in the following subsections only one experiment conducted with a MuSA is described; all the results obtained from the other experiments were consistent with those

reported below. These experiments helped to determine which configuration is more stable and reliable between *Differential, Single-Ended Reference and Single-Ended Screw.*

In the following figures, only data from the large electrodes are reported, since the small ones perfectly followed the same oscillations and did not add any new information to the figures.

3.1.1.1 Mean trials in time and time-frequency domain

The first analytical step of the setup validation was the computation of the mean trial to detect possible artefacts or disadvantages due to an erroneous connection between reference and ground.

The plots reported in Figure 13 and Figure 14 represent the mean trial obtained from the average of 100 trials, previously band-pass filtered at 70-300 Hz for each recording configuration tested: Differential (A), Single-Ended Reference (B), Single-Ended Screw (C).

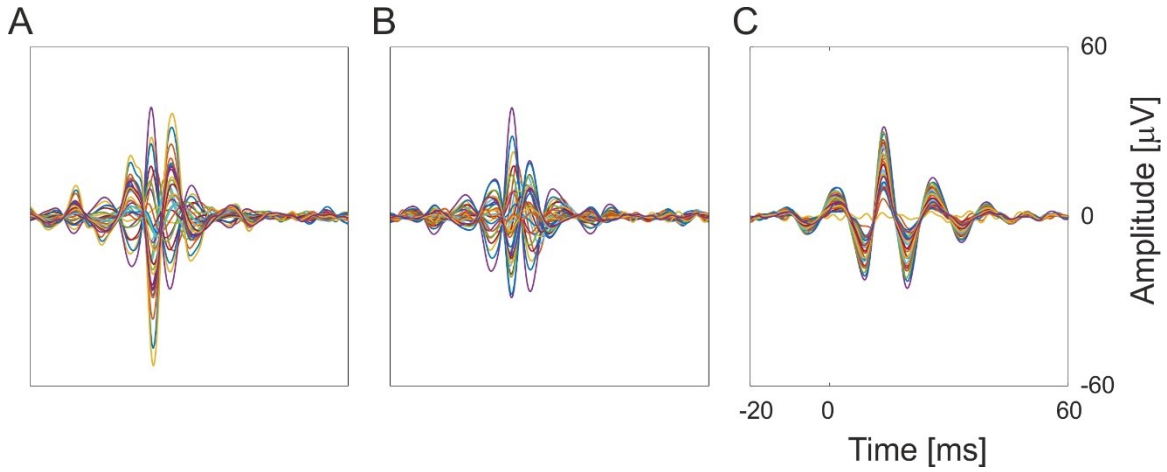


Figure 13. Mean SEP trial for Differential, Single-Ended Reference and Single-Ended Screw configurations. Each plot represents the average of 100 trials for all the recording electrodes, reported for Differential (A), Single-Ended Reference (B) and Single-Ended Screw (C) configurations. The recordings were performed with the MuSA.

From the data visualization in Figure 13, there is an apparent difference across the three recording approaches. In particular, we assist to sign-inverted activity for some electrodes during the recording with Differential (A) and Single-Ended Reference (B) configurations,

but not with Single-Ended Screw (C). The details of this phenomenon are more evident in Figure 14. The Single-Ended Screw (C) configuration is characterized by a waveform propagating from the top right to the bottom left, with a gradual reduction of the evoked amplitude. In contrast, in both Differential and Single-Ended Reference configuration recordings, the signal starts propagating from the top right, but shows a fast reduction in the peak amplitude and a sudden reversal of the sign in the middle of the recording area (marked with a red line). This indicates that the reference onsite has a strong effect on both Differential and Single-Ended Reference configurations. To understand the effect in the time-frequency domain, the mean spectrograms were computed in the same frequency band (see Figure 15A-B). The main effect consisted of delocalized power due to the presence of the reference. Moreover, in both Differential and Single-Ended Reference configurations, the spectrogram of electrode 4 appeared to be less powerful compared to the Single-Ended Screw configuration. A similar behavior is also found in the electrodes in the central portion of the device.

This evidence demonstrates that the reference onsite can significantly change the spatial localization as well as the sign and amplitude of the signals, depending on what the reference is recording and whether it is connected to the reference pin (Differential) or to both reference and ground pins (Single-Ended Reference).

3.1.1.2 Correlation analysis

The mean correlation maps provide additional evidence that the Single-Ended Screw configuration is the optimal solution for recording with micro-ECoG devices, particularly when the onsite reference is too close to the source. The correlation across channels changes sign when the reference is connected either to the reference pin or to the ground pin; thus, signals are negatively correlated (see Figure 16A-B). In contrast, the correlation between those same electrodes is strictly positive when the reference and the ground are connected to the skull screw (C).

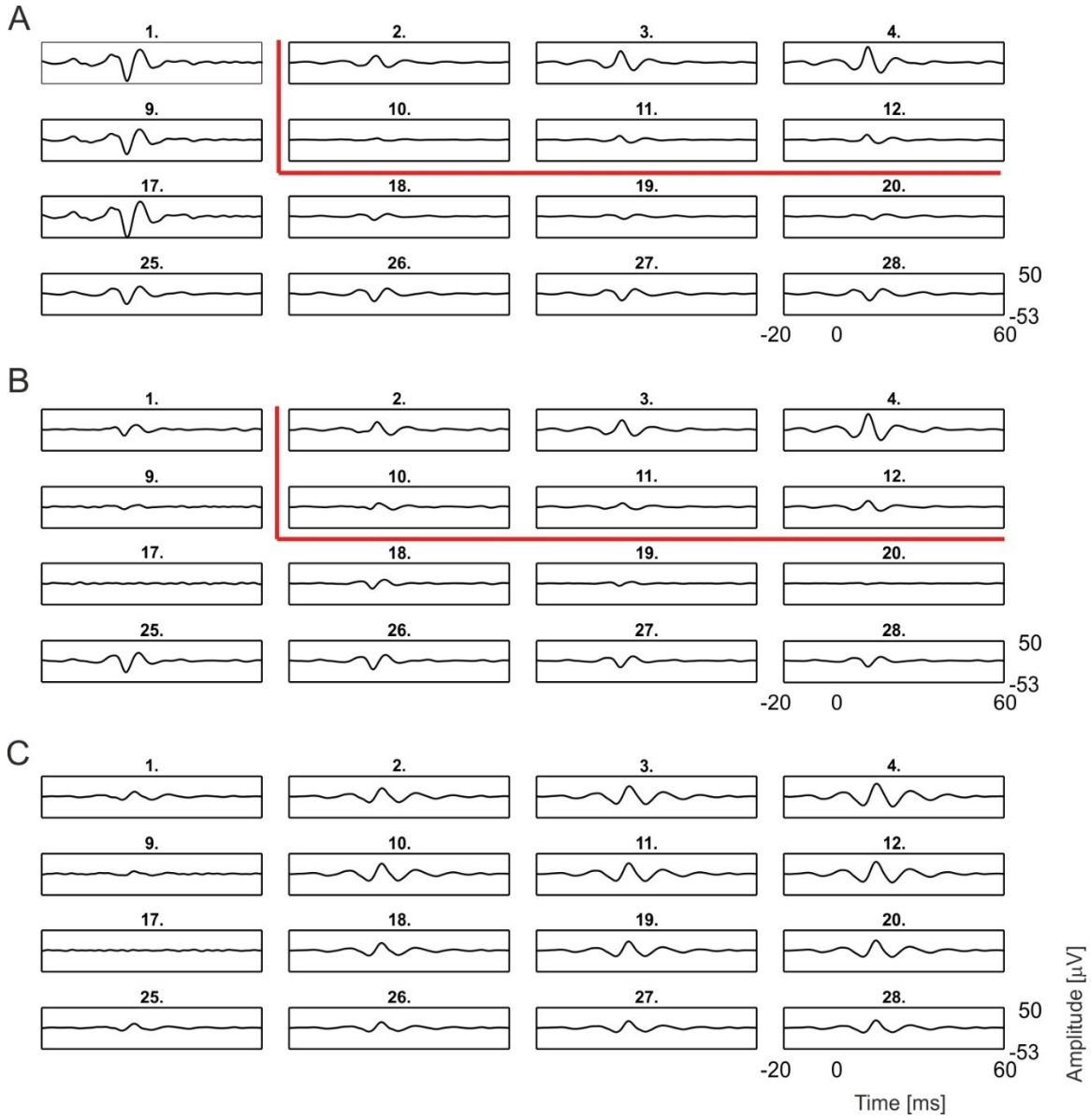


Figure 14. Mean SEP trial for Differential (A), Single-Ended Reference (B) and Single-Ended Screw (C) configurations. Each plot represents the average of 100 trials. For better clarity, only data from the large electrodes are reported, since the small ones behaved identically. The red line indicates the reversal of the peak due to the presence of the onsite reference (A) and (B), which was not present when the reference was connected to the skull screw (C).

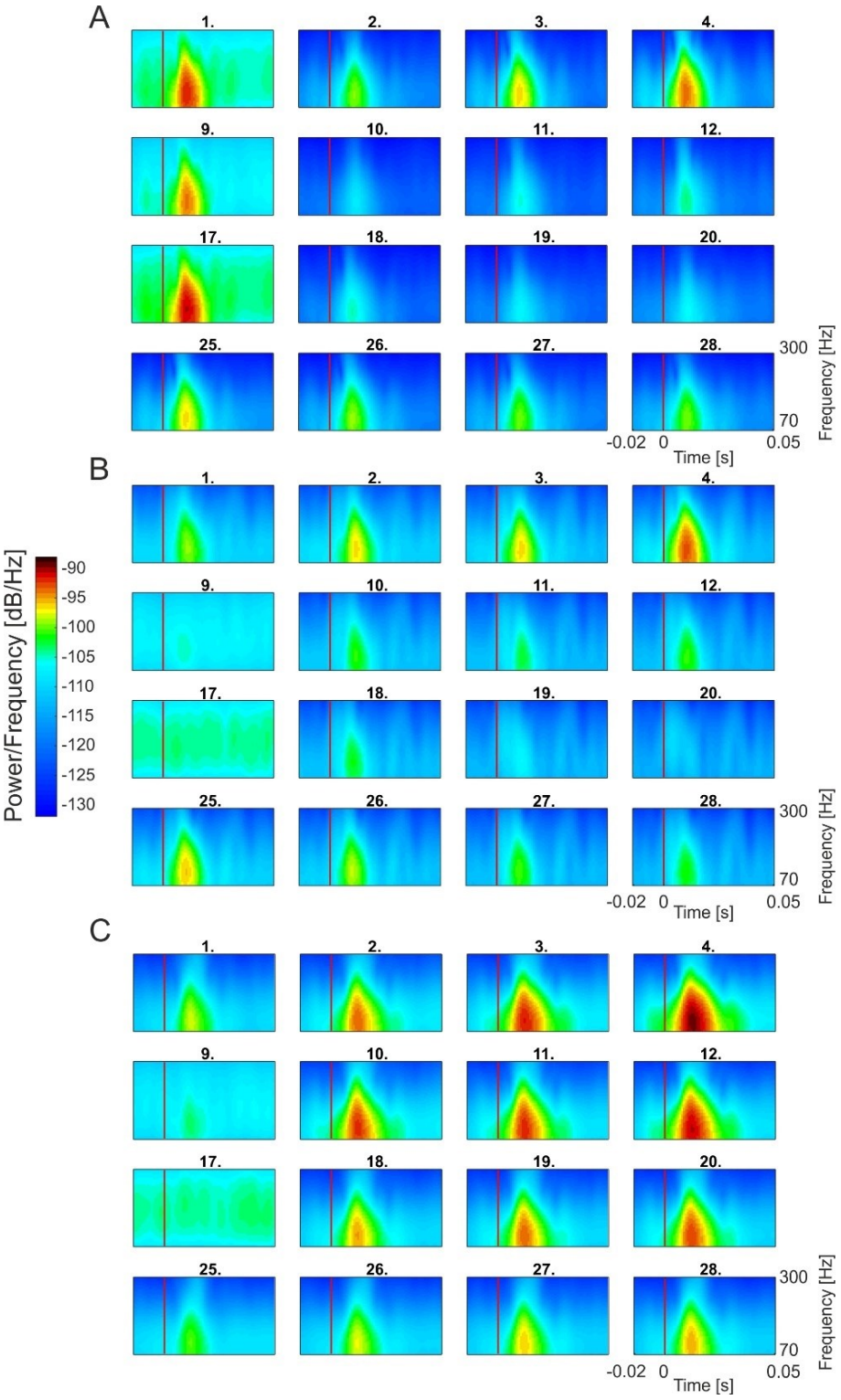


Figure 15. Mean spectrograms for Differential (A), Single-Ended Reference (B) and Single-Ended Screw (C) configurations. Each plot represents the average of 100 trials. Again, for clarity, only the large electrodes are reported (see labels). The vertical red line represents the timing of the stimulus.

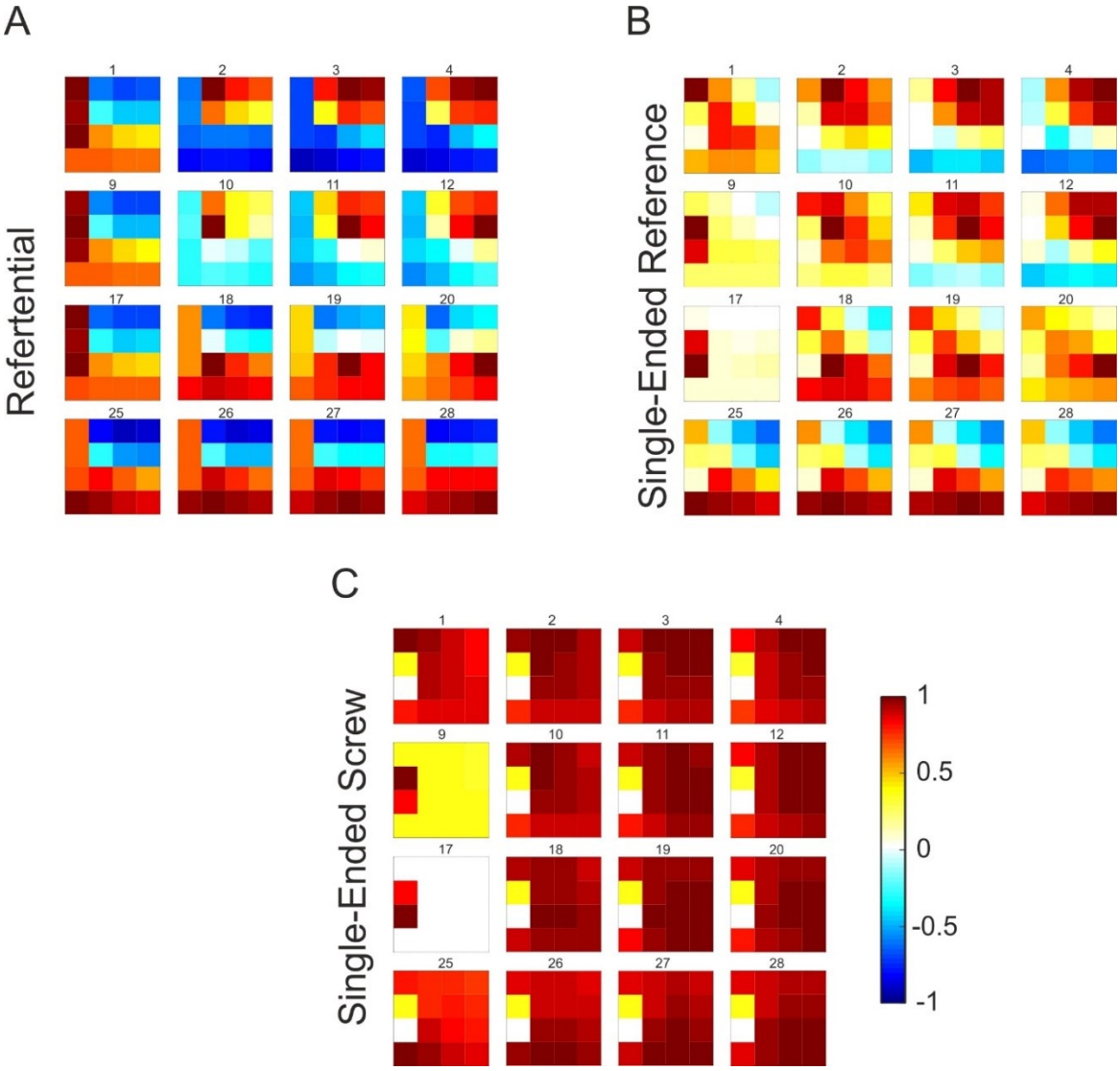


Figure 16 Mean correlation maps obtained for the three recording configurations: Differential (A), Single-Ended Reference (B) and Single-Ended Screw (C). Each labelled square represents one of the large channels (see labels) and contains the correlation coefficients for that channel against all the others. The maps share the same color-scale, from -1 (anti-correlated) to 1 (correlated) signals.

3.1.1.3 Interpolated maps from peak-to-peak amplitude and evoked-to-spontaneous ratio

The maps computed for each setup configuration are reported in Figure 17: Differential configuration (A), Single-Ended Reference (B) and Single-Ended Screw (C). Since the device maintained the same position over the barrel cortex for the entire experiment, the change in the median peak-to-peak amplitudes can be due to either the recording configuration

or the level of anesthesia. However, it should also be considered that the latter factor would reasonably impact all the electrodes equally.

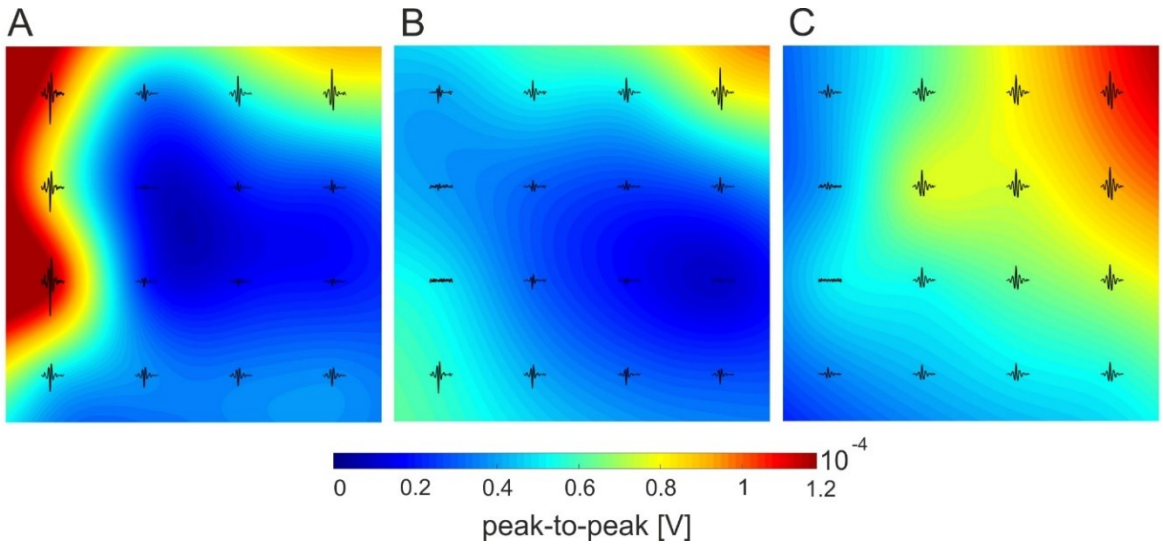


Figure 17. Bidimensional interpolated maps obtained for the three different configurations: Differential (A), Single-Ended Reference (B), Single-Ended Screw (C). The maps were computed starting from the median peak-to-peak amplitude for the three different recordings. The device was maintained in the same position during the entire experiment. Blue denotes the lowest values, and red indicates the highest values. The recordings were performed with the MuSA, and only the large channels are reported.

The interpolated maps are a useful tool to locate the source of the signal, but they do not take into account the background noise; thus, they cannot quantify the electrode's performance. As previously mentioned, a straightforward measure to represent the electrode's performance while recording evoked potentials is the evoked-to-spontaneous ratio (ESR). This measure is not based on the mean peak-to-peak amplitude but also on the background noise. The ESRs for the three different recording configurations performed with the MuSA are reported in Figure 18. When observing the mean trial and the interpolated map related to the Differential Configuration (see Figure 14A), channel 17 appear to record perfectly the SEPs. However, the ESR values were considerably low (Figure 18A) because of the high background noise, visible from the spectrograms (Figure 15).

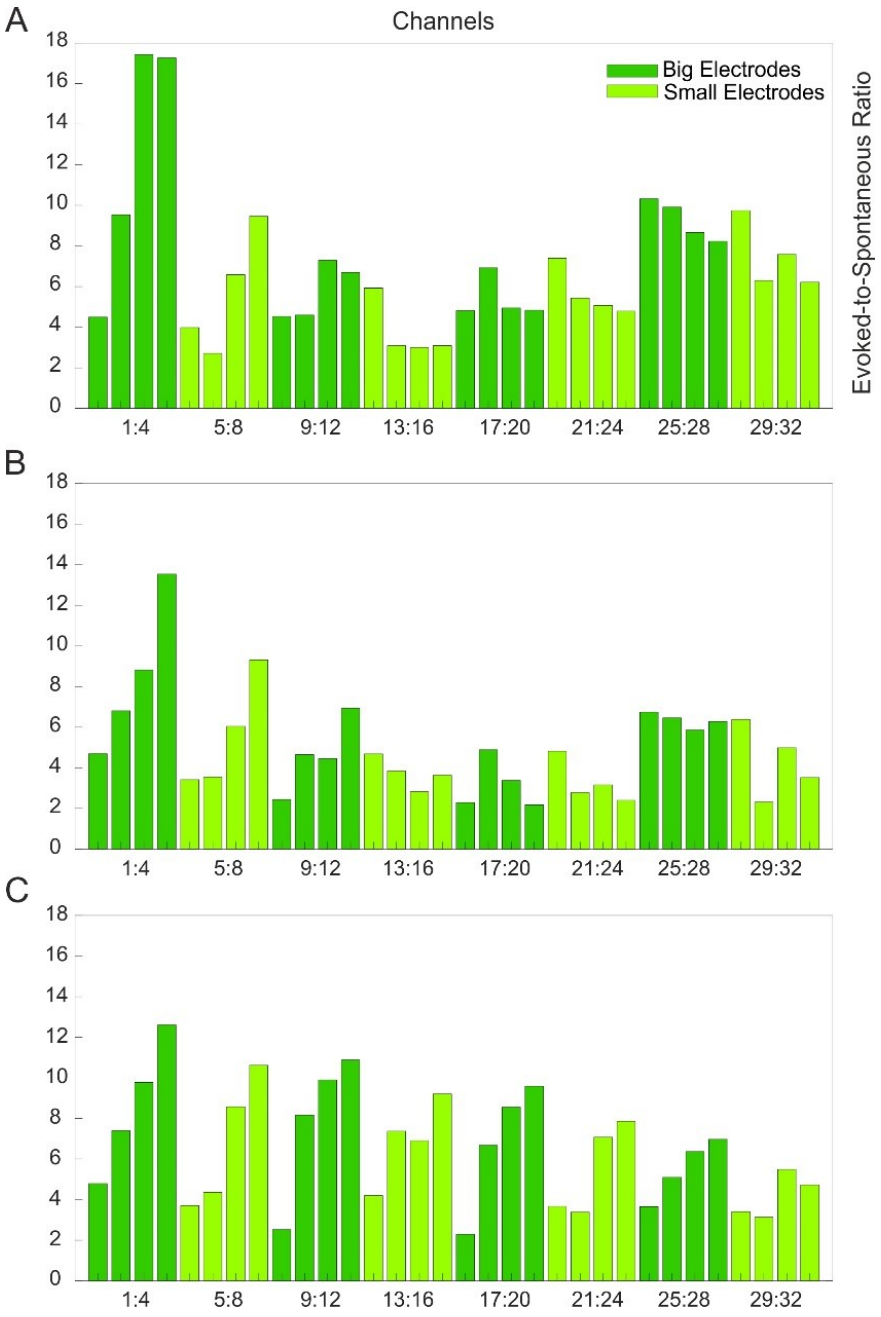


Figure 18. Evoked-to-spontaneous Ratio for the three setup configurations tested with the MuSA. The small electrodes, in light green, usually performed slightly worse than the corresponding large electrodes (in dark green); this behavior is due to the difference in the impedance of the two groups. The ESR ratios reflect the behavior observed with the spectrograms for the three configurations. The electrodes with the lowest ESR in Differential (A) and Single-Ended Reference (B) configurations show a higher ratio than the Single-Ended Screw (C).

The ESR is an important measure for an additional reason: this measure also quantifies differences across experimental conditions and, in this particular case, across the three configurations. In Differential (Figure 18A) and Single-Ended Reference (Figure 18B) configurations, the ESRs of channels 19 and 20 were considerably low when compared to the Single-Ended Screw values.

3.1.2 Discussion

To understand the most appropriate configuration of ground and reference for micro-ECoG recordings, several methods have been developed. These analyses provide evidence that the three tested configurations (Single-Ended Reference, Single-Ended Screw and Differential) record very different information and thus are not exchangeable. The main difference between the configurations can be attributed to the onsite reference, which can cause major effects in both Differential and Single-Ended Reference configurations: the activity detected by the reference affects the signal recorded with active electrodes. Therefore, the reference should not be on top of the signal source of interest. Additionally, depending on the similarity between the onsite reference and recording electrodes oscillations, we can deduce different implications. Specifically, if reference and recording electrodes show high positively correlated signals, there will be almost no difference between their oscillations and, thus, lower activity or “inverted” activity will be recorded depending on which signal is stronger. This inversion was confirmed by the correlation analysis results. Here, I show that the reference can nullify, amplify or invert signals, creating negative correlations in case of inverted signals (see Figure 16A-B). The second issue impacts the Single-Ended Reference configuration. In this case, the ground pin and the reference pin are shorted to the reference onsite, which fluctuates with the potential of local brain electrical activity, causing fluctuations in the ground potential as well. This configuration is usually not indicated because it can be a source of noise. In contrast, in the Differential configuration the ground pin is connected to an external screw implanted in the skull of the rat, which in turn is in a Faraday cage connected to earth ground. In this case, the potential is highly stable, and it does not affect the recorded signal.

Since it is also possible to reference the data offline by choosing a recording electrode as reference, the Single-Ended Screw configuration may be considered to be the most reliable solution. However, the reference could be not recorded in Differential or Single-Ended Reference because of setup constraints; thus, the data could not be transformed offline by re-referencing. This preliminary study was conducted both with the MuSA and with the Epi device. Since the results for the two devices were coherent, only the findings using the MuSA were reported. These findings were fundamental to choosing the most reliable recording configuration, employed for the following experiments.

3.2 Contribution to MuSA validation

3.2.1 Results

Once the setup was validated, further experiments were conducted to test the MuSA devices in the best possible recording configuration. The devices have been tested both acutely implants to determine their recording reliability for high frequency components. In this project, I was responsible for data pre-processing and visualization for testing the signal quality and possible differences between large and small electrodes.

In the following subsections, I focus on the results obtained in the analyses of two MuSA devices, from now on *device1* and *device2,* acutely tested in the same rat. Each device was tested in three different positions over the barrel cortex.

3.2.1.1 Time-frequency analysis

The average spectrograms obtained for all the large channels of both devices are reported in Figure 19. The results of *device1* spectrograms are shown in A-C for three recording sessions and different device positions over the rat barrel cortex. The results for *device2* are reported in D-F, also for three distinct positions.

3.2.1.2 Interpolated maps from peak-to-peak amplitude

The interpolated maps reflect the different spatial localization of the device above the barrel cortex (see Figure 20). The maps were computed as usual starting from the median peak-to-peak amplitudes for the different recordings, i.e., six different positions recorded

with two devices above the barrel cortex. The peak-to-peak amplitudes were calculated starting from the band-pass filtered signals in the range 200-1000 Hz. This approach was important to demonstrate that high-frequency activity recorded with micro-ECoG electrodes propagates and spreads across the cortex surface. Indeed, very close electrodes can record the SEP-related activity with different amplitudes, even during multi-whiskers stimulation within a small recording area.

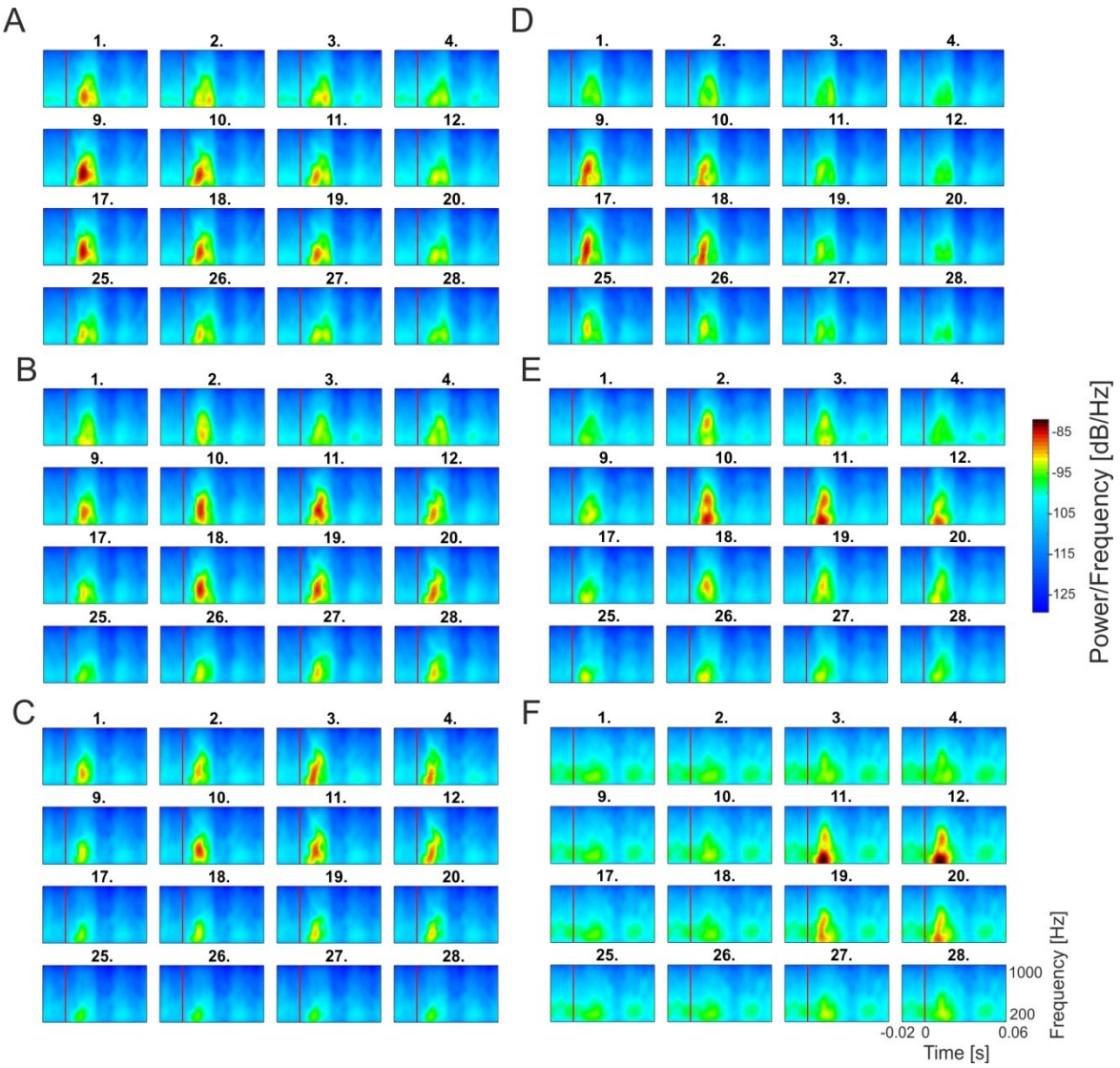


Figure 19. Average spectrograms obtained for the sixteen large electrodes in the MuSA devices, one for each recording; A-C for device1, and D-F for device2. The plots share the same color-scale reported with the color-bar. On the x-axis, the time is reported in seconds, while the red vertical line corresponds to the stimulation onset. The y-axis represents the frequency range (from 200 to 1000 Hz).

3.2.1.3 Evoked-to-spontaneous ratio

The evoked-to-spontaneous ratios for the six different recordings performed with the MuSA are reported in Figure 21. The ESR reflects a change in the array positions with respect to the source of the evoked activity during the different recordings, as well as the quality and capacity of the recording electrodes. The highest ESR values of almost every recording derive from the large electrodes, while the small contacts usually perform slightly worse than the corresponding large ones. The results of the one-way analysis of variance (ANOVA) were obtained by comparing the ESR values of each large electrode to the corresponding small one independently for each recording session (see Figure 22). This analysis provides further evidence that large electrodes produce higher quality recordings of high frequency components. This behavior might be due to the difference in the impedances of the two groups of electrodes, as described in a previous study [86].

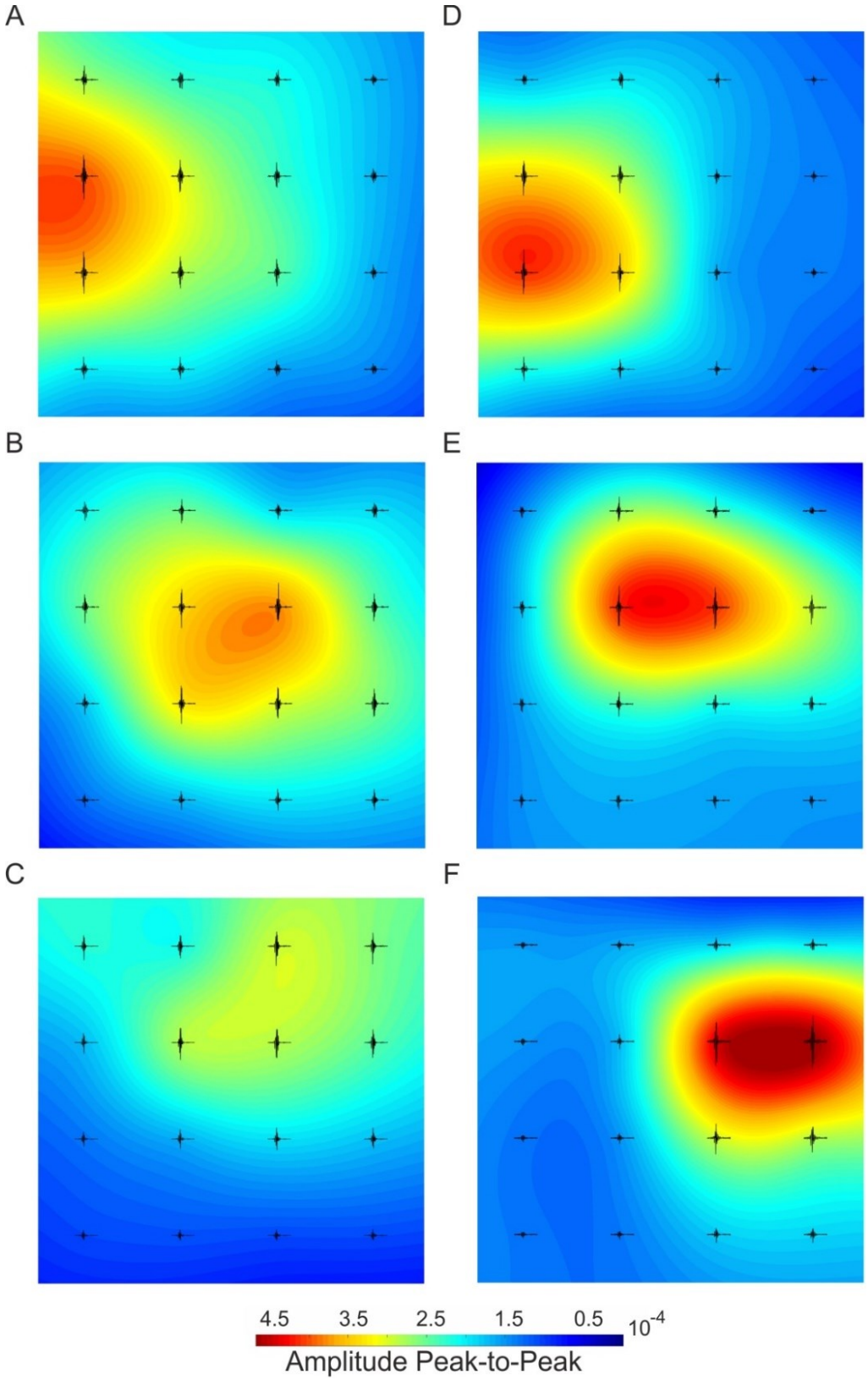


Figure 20. Bidimensional interpolated maps obtained for the recordings performed with the MuSA devices. The maps were computed starting from the median peak-to-peak amplitude for the six different recordings acutely performed with two devices, (A-C) device1 and (D-F) device2. The color-

code indicates blue for the lowest values and red for the highest values. The plots share the same color scale, reported in the bottom with a color bar.

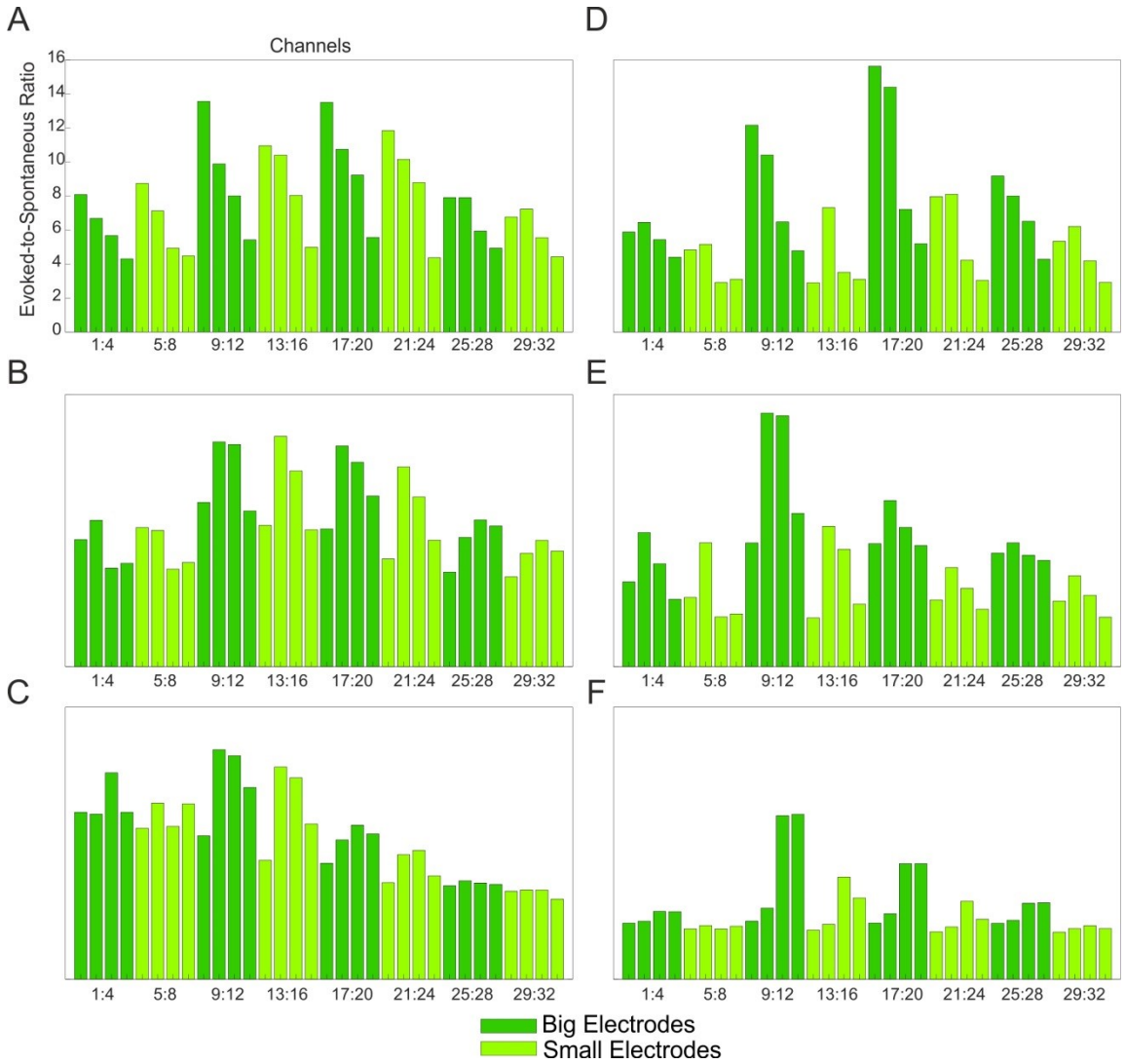


Figure 21. Evoked-to-spontaneous ratio for the six different recordings performed with the MuSA devices. The ratio accurately reflects the change of the array positions with respect to the source of interest; the electrodes with highest ESR change for each position. The small electrodes, in light green, usually perform slightly worse than the corresponding large electrodes, shown in dark green.

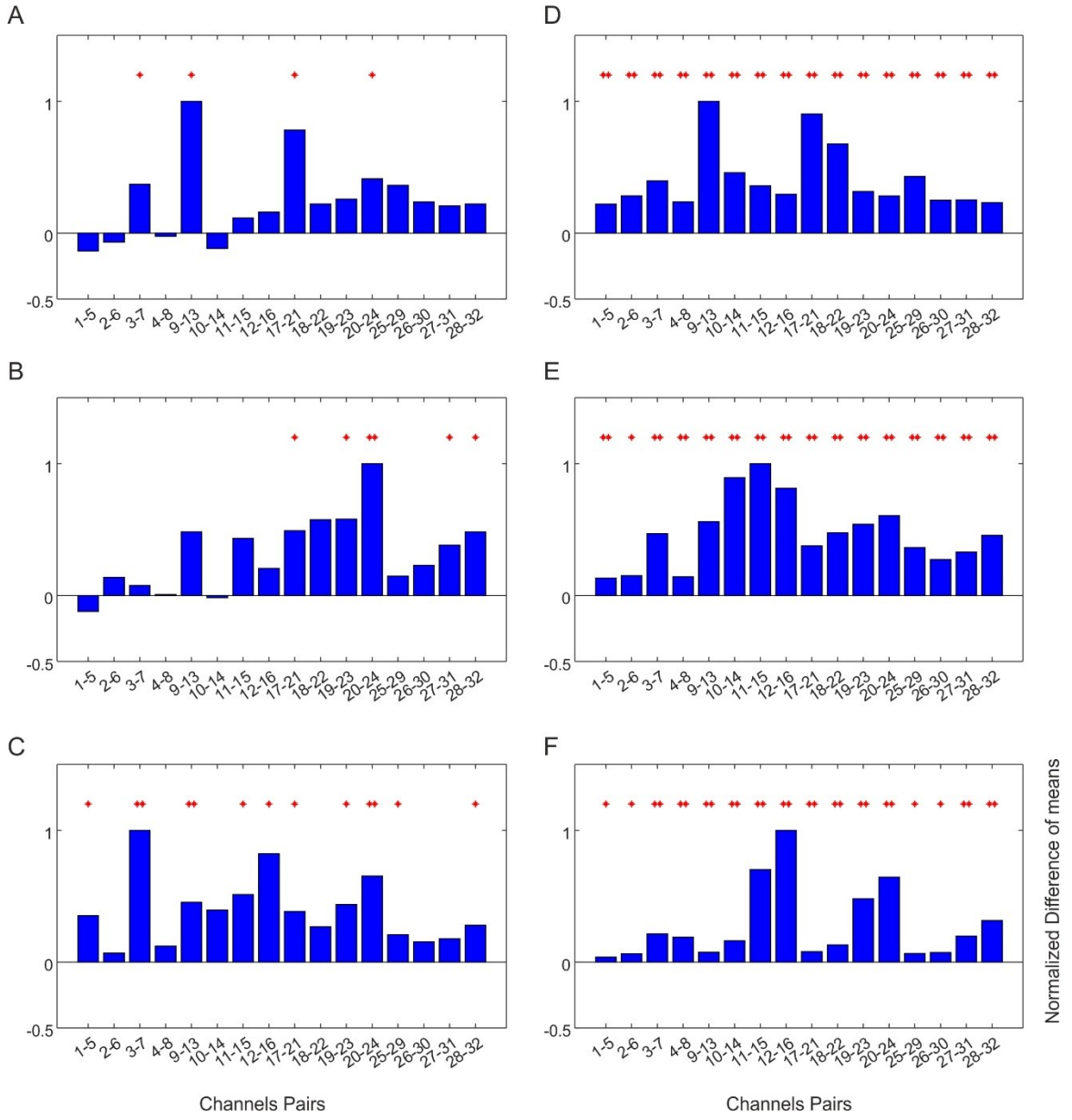


Figure 22. One-way ANOVA results comparing the ESR values for large and small electrodes. The bar plots show the difference between the estimated group means, normalized by the maximum value for each pair of electrodes (large minus the respective small). The group means are significantly different from each other if one asterisk ($p<0.05$) or two asterisks ($p<0.01$) are reported above.

3.2.2 Discussion

The main goal of this study was to validate the MuSA, a new highly conformable Polyimide-based device with platinum electrodes. This device was validated during several *in* vivo experiments performed by a biologist. The experiments consisted of peripheral

mechanical stimulation of the whiskers with simultaneous recording of barrel cortex activation in rats. My role was to develop the scripts in order to visualize the data and measure the quality of the micro-ECoG recordings. The analyses provided evidence that the MuSA devices can record high frequency multi-unit signals, and electrode capability is not dramatically affected by the size of the electrodes. These results confirm that high frequency components (above 200 Hz) of neuronal signals can be recorded directly from the cortex with a spatial resolution never reached before with cortical recordings. Indeed, the previously-attained spatial resolution reached until now is 1000 µm, and only 2000 µm for speech investigations [83, 96]. Such a minimally-invasive and ultra-conformable device could be helpful in both research and clinical applications, from understanding physiological network activity to BCI research [29, 42, 79, 80].

3.2.3 Related Publications

[86] M. Vomero, M.F.P. Cruz, E. Zucchini, A. Shabanian, E. Delfino, S. Carli, L. Fadiga, D. Ricci, T. Stieglitz, Achieving ultra-conformability with polyimide-based ECoG arrays, 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 4464-4467.

3.3 Contribution to CaLEAF validation

The CaLEAF electrodes validation was conducted in five experimental sessions with different devices. In the following subsections, only one experiment will be described, since all the obtained results were consistent with those reported below [87].

The device was tested first by stimulating all the whiskers and then by stimulating only single whiskers to test whether all the electrode diameters were able to detect and discriminate the evoked responses from the background noise.

3.3.1 Results

3.3.1.1 Time-frequency analysis

To determine whether the size of the electrodes was an issue in recordings of high-frequency components, time-frequency analysis was conducted on the signals, which were band-pass filtered at 200-1000 Hz. Apart from one electrode with a diameter of 300 µm, all the other electrodes were able to record the signals of interest during multi-whiskers stimulations (see Figure 23A-C). However, the smaller electrodes showed higher background noise compared to the larger electrodes. This evidence was also confirmed by the spectrograms obtained for the single-whisker stimulations (Figure 24). Indeed, the smallest electrodes (5-8) showed the most powerful spectrograms but also the highest background noise.

3.3.1.2 Evoked-to-spontaneous ratio

The ESR results are reported in Figure 25 for both the multi -whiskers stimulation performed testing three position (indicated as P1-P2-P3) of the ECoG device over the barrel cortex (A-C), and the single-whisker stimulations, C2 (B), D4 (C) and B2 (D). The ratio values reflect both the change in the array positions with respect to the source of interest and the different types of stimulation, multi -whiskers (A-C) and single-whisker (D-F).

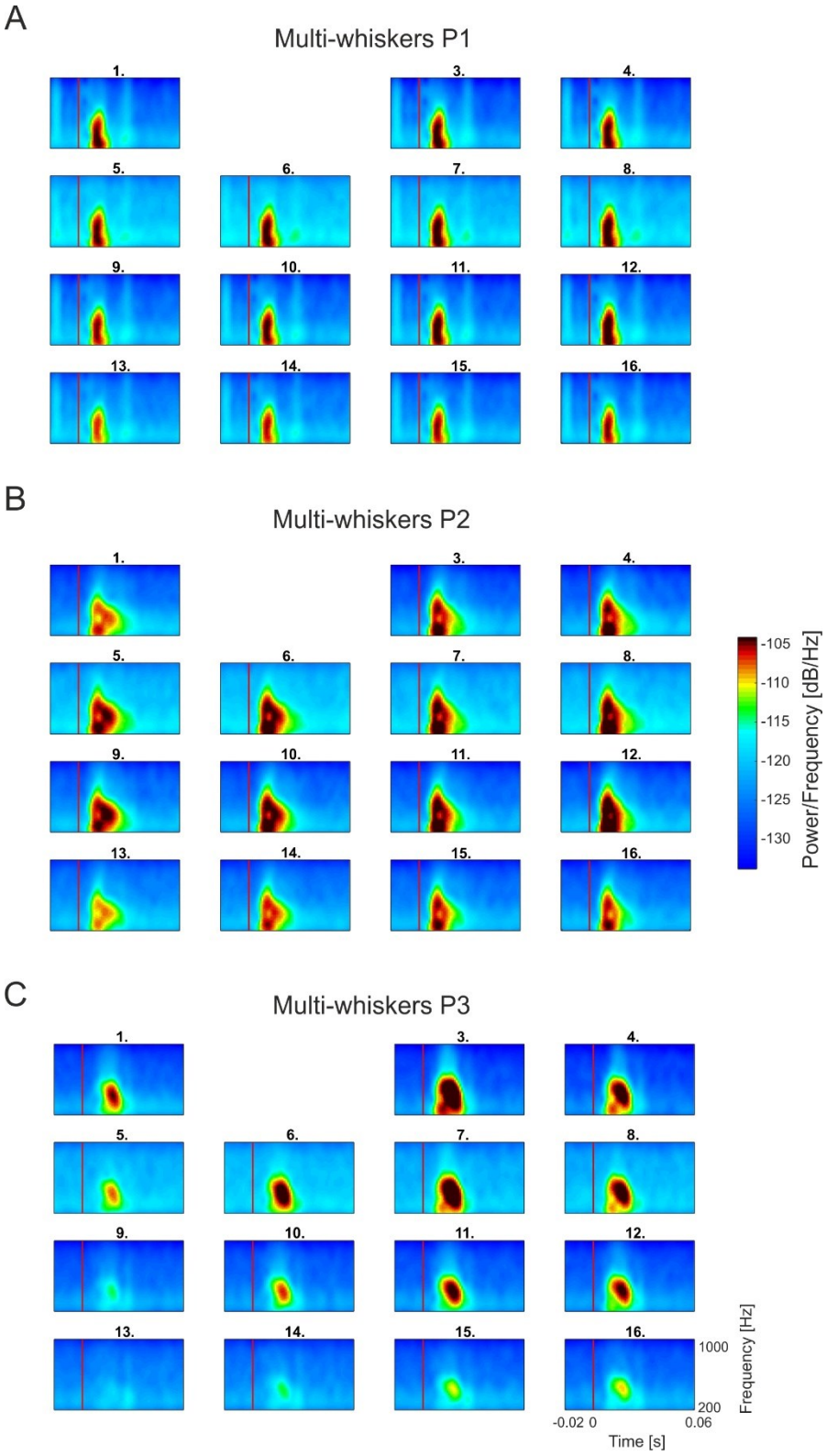


Figure 23. Mean spectrograms for the CaLEAF device obtained by averaging 100 trials for three different recording sites during multi-whiskers stimulations. Electrode 2 was not working properly so this position was left empty to avoid biasing the color scale. Electrodes sizes: 1-4 300 µm; 5-8 50 µm; 9-12 200 µm; 13-16 100 µm.

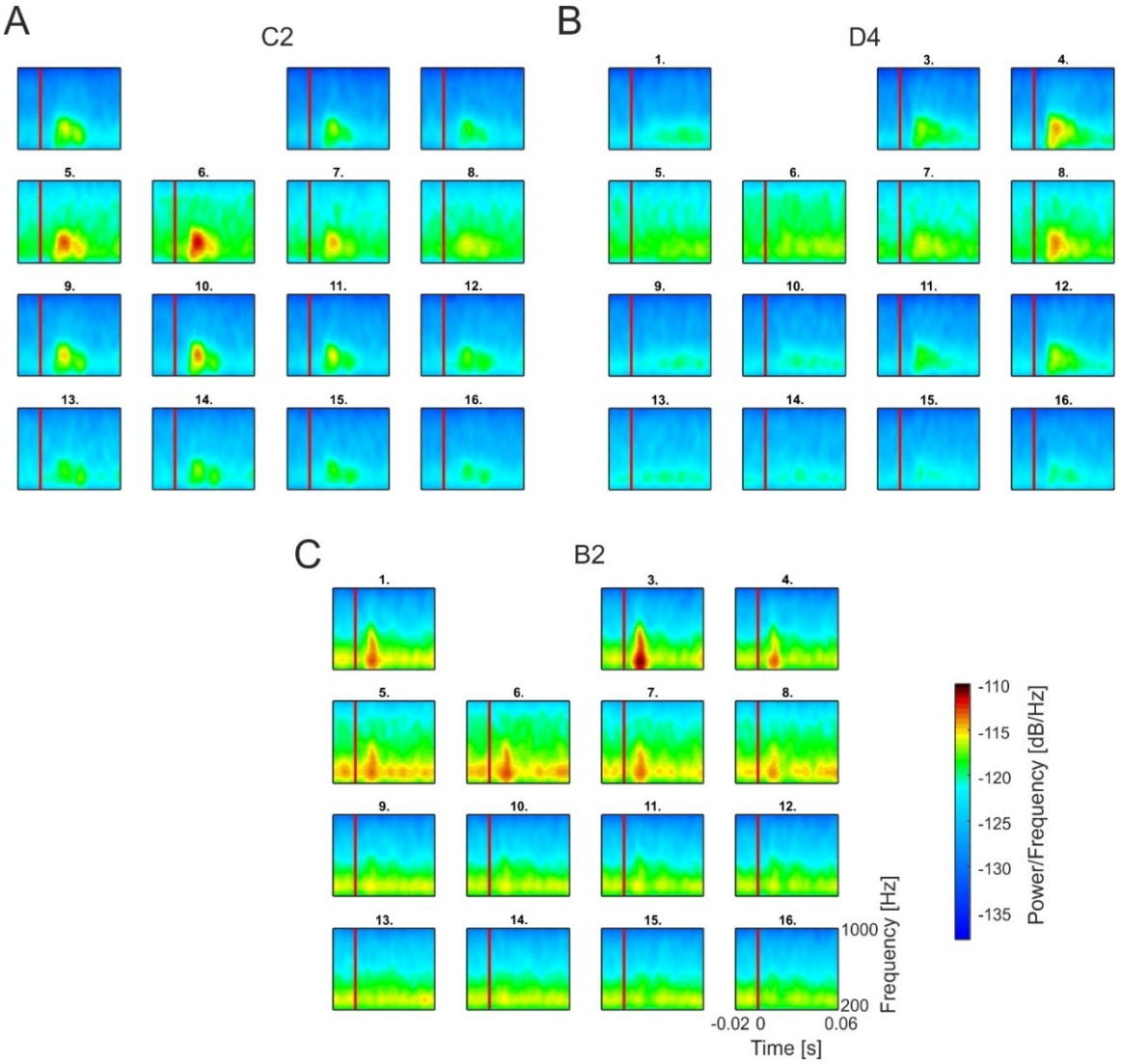


Figure 24. Mean spectrograms for the CaLEAF device, obtained by averaging 100 trials of three different single-whisker stimulations, (A) C2 stimulation (B) D4 stimulation (C) B2 stimulation. Each single square represents one electrode (see label), while the red vertical line corresponds to the stimulation time. All the plots use the same color-scale.

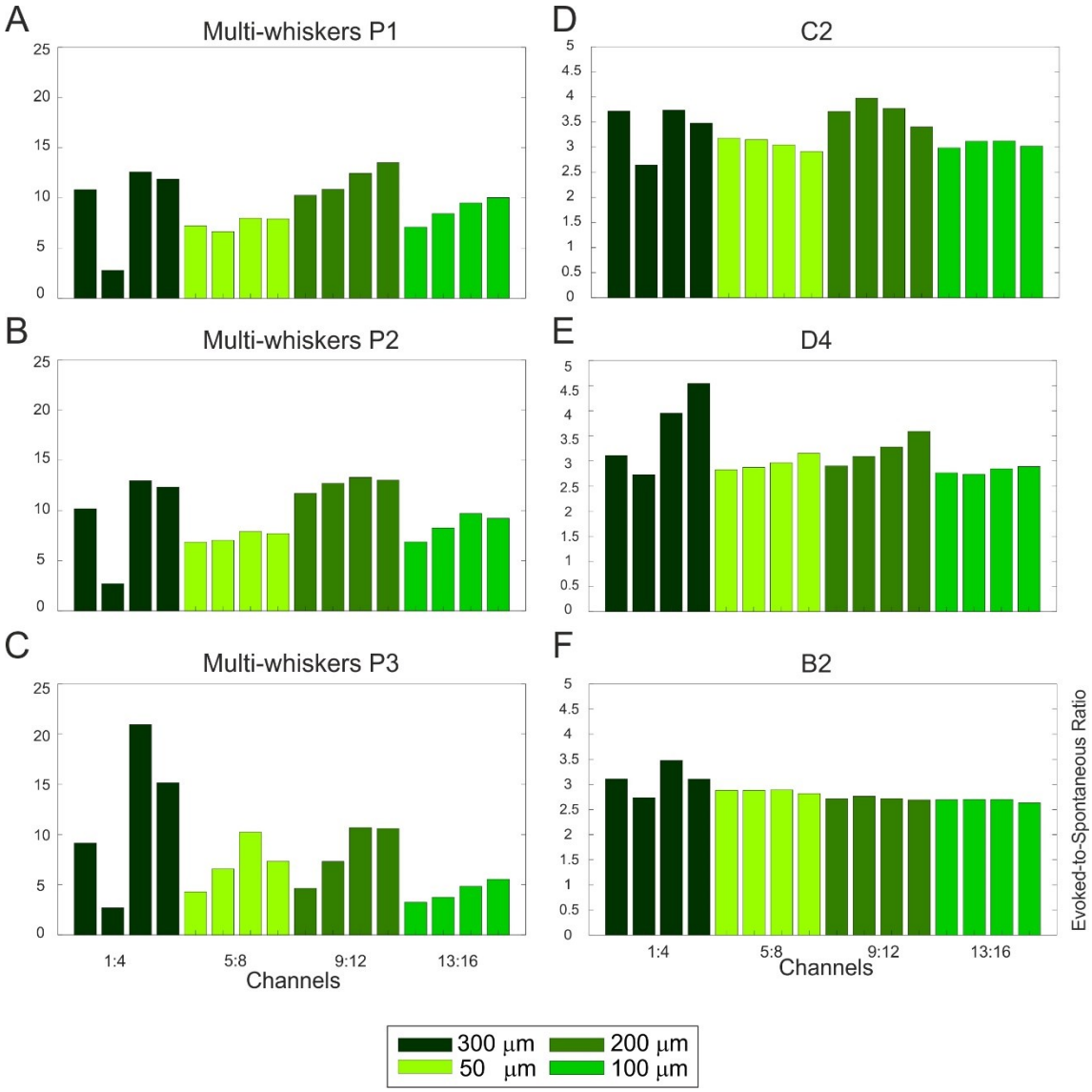


Figure 25. Evoked-to-spontaneous ratio for the different recordings performed with the CaLEAF device. The ratio reflects both the change in the array positions with respect to the source of interest and the different types of stimulation, multi-whiskers (A-C) and single-whisker (D-F). The amplitude scale for the multi-whiskers stimulations (on the left column A-C) is different from the one set for single-whisker stimulation (D-F) in order to better understand the behavior of the electrodes.

3.3.2 Discussion

The main goal of this study was to validate another new, highly conformable polyimide-based device with glassy carbon electrodes, the CaLEAF. Glassy carbon has high potential in neural applications since it is biocompatible and electrochemically inert, and is capable

of simultaneously serving as a stimulating and recording site [87]. The device was validated during several *in vivo* experiments performed by a biologist. The experiments consisted of peripheral mechanical stimulation of the whiskers while recording the activity elicited in the barrel cortex of rats. My role was to analyze the data, adapting the scripts already developed for the MuSA validation. The analyses conducted on this dataset demonstrated that all the electrodes could record the high frequency components of the SEPs, even though small electrodes showed higher background noise than large electrodes. Excluding a broken contact (2), electrodes with larger diameters (300 µm) showed the best performance, while background noise was evident in both spectrograms and ESRs of the smallest electrodes (see Figure 23). Therefore, the relevant difference between the channel performances seemed to be related to the different electrode diameters. Nevertheless, the ESR values of the smallest electrodes were not different from electrodes with 200 µm diameter (see electrodes 9-12 in C and F). However, when using the ESR to evaluate electrode performance, it is important to note that the values depend on electrode dimensions as well as the distance of the electrode from the signal source location. Since it is not possible to make a direct comparison between close electrodes of different sizes in this device, the performance of these electrodes cannot be statistically compared. Finally, when only single whiskers are stimulated, smaller SEP amplitudes and lower ESR values were reported for all the electrodes, which was expected given the reduced population of neurons involved in the task (Figure 25).

3.3.3 Related Publications

[87] M. Vomero, E. Zucchini, E. Delfino, C. Gueli, N. Mondragon, S. Carli, L. Fadiga, T. Stieglitz, Glassy Carbon Electrocorticography Electrodes on Ultra-Thin and Finger-Like Polyimide Substrate: Performance Evaluation Based on Different Electrode Diameters, Materials, 11 (2018) 2486.

3.4 Translational application in humans: predicting speech-related motor preparation

In this section, I report data recorded from the speech arrest area during a naming task. After recording the brain activity during speech production, several analyses and methods were applied offline to visualize and investigate the spatiotemporal relationships between signals recorded with sub-millimeter-spaced microelectrodes. I provide evidence that µECoG recordings contain the necessary information to predict when the subject is going to speak, using two different devices.

3.4.1 Results

3.4.1.1 Data visual inspection

Segmentation and band-pass filtering

The recorded signals (see Figure 26) were filtered using Butterworth band-pass filters of order 8 in the traditional beta (15 - 30 Hz), gamma (30 - 60 Hz) and high-gamma (70 - 150 Hz) bands, applying the function *filtfilt()* to avoid phase distortion; 50 Hz and the harmonics up to 200 Hz were removed with notch filters. The timing of the speech production was used to segment the filtered data into trials. For visual inspection purposes, the trials were segmented from 500 ms before to 500 ms after the speech onset. The resulting band-pass filtered signals of one trial are reported for some selected channels in Figure 27B-D as well as the raw signals (A).

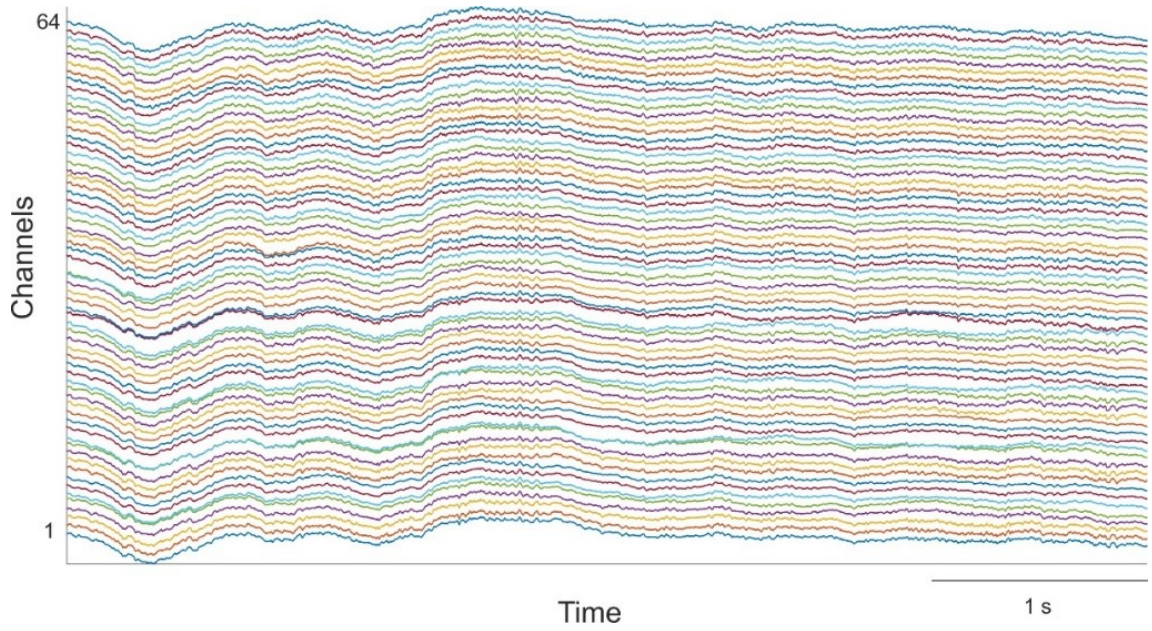


Figure 26. Raw µECoG signals extracted from the recording performed with the Epi device. Channels on the y-axis extend from 1 (bottom) to 64 (top), and the time interval is 5 seconds long. All the electrodes were recording properly during the experiment.

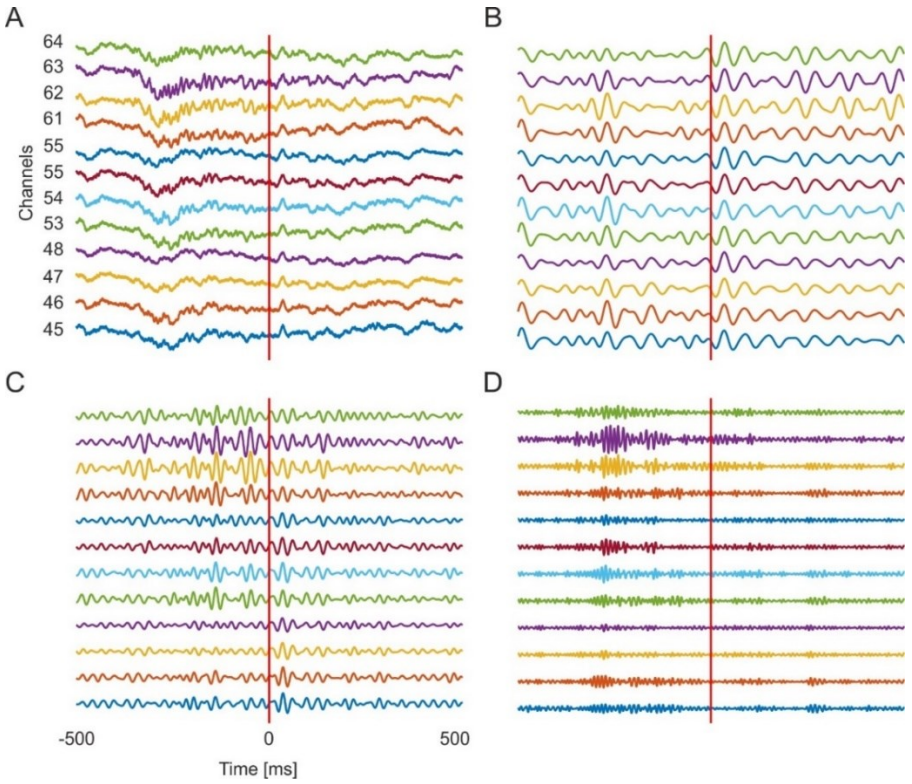


Figure 27. Raw and band-pass filtered signals aligned to the voice onset for one trial recorded with the Epi device. A) The raw signals extracted from an interval lasting approximately 1 second are reported for 12 selected channels (see y-label). For the same channels and the same time interval,

the band-pass filtered signals in 15-30, 30-60 and 70-150 Hz are reported (B-C-D, respectively). The vertical red line indicates the timing of the speech onset. The amplitude scales are independent for each plot to allow a better visualization.

Correlation analysis

The correlation analysis was conducted in two different steps to inspect possible patterns across the different frequency bands. The first informative result was derived from the correlation profile analysis (see Figure 28A). Here, the mean correlation coefficient obtained by averaging across trials and electrodes of same distance is shown for the inspected frequency bands. Consistent with previous literature [77], the three computed profiles show decreasing trends, with higher values for lower frequencies; interestingly, the high-gamma correlations show a consistent drop (50% reduction) for inter-electrode distances in the range from 0.6 mm to 2.4 mm, indicating that electrodes with a distance greater than 2.4 mm between each other are not highly correlated during the task. Additional evidence is reported in Figure 28D, where the spatial distribution of the correlation coefficients averaged across trials is represented according to the electrode position on the array. As illustrated in the figure, the high-gamma correlations between the most active channels (bottom-right portion of the array) are locally high, but they steeply decreased to 50% of their initial values within few electrodes.

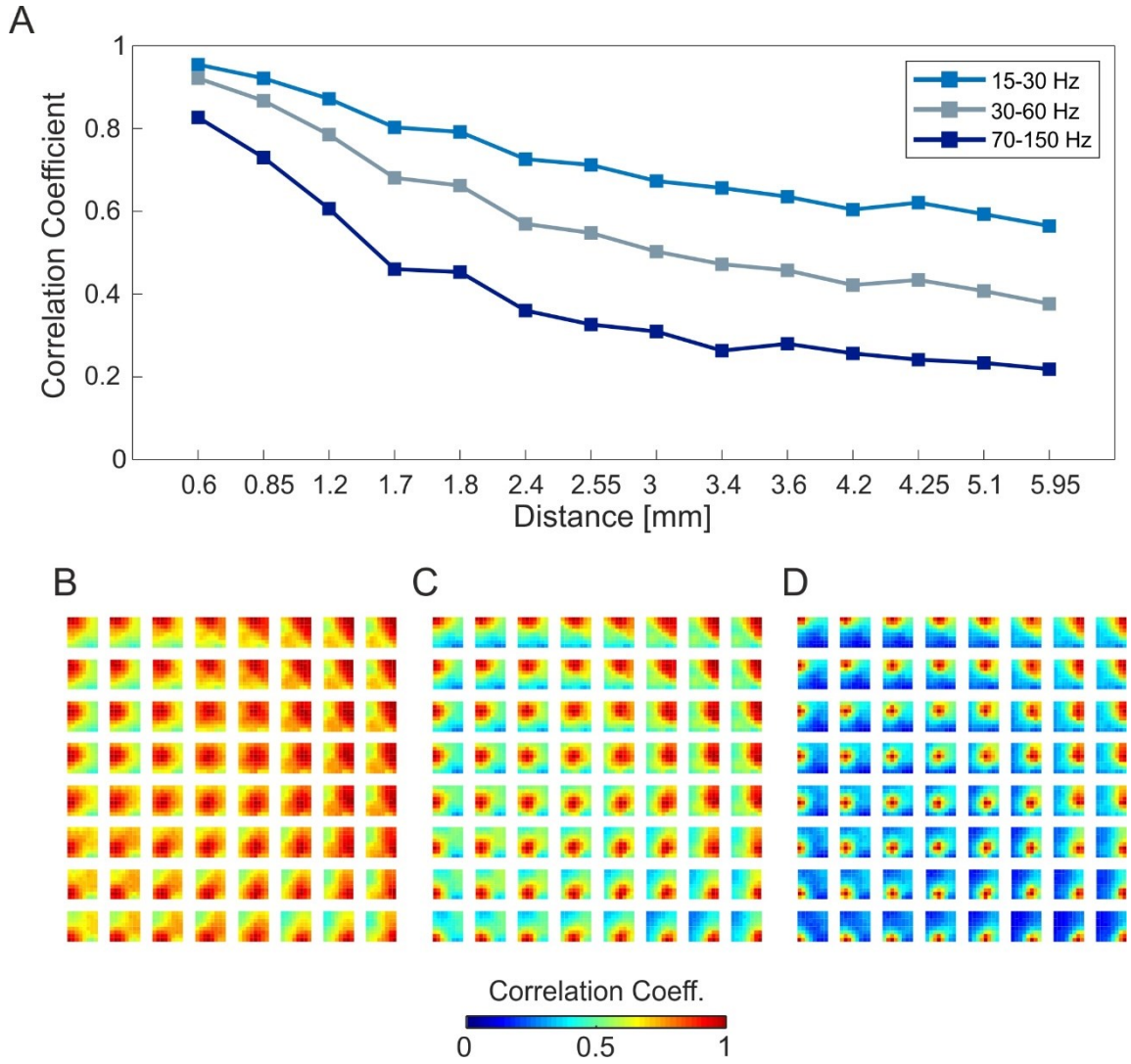


Figure 28. A) Correlation profiles obtained considering signals from equidistant electrode pairs and averaging across trials. Colors represent different frequency bands. Mean correlation maps were obtained averaging single-trial correlations, computed for signals band-pass filtered in 15-30 Hz (B), 30-60 Hz (C) and 70-150 Hz (D). Each matrix of the plot represents the correlation coefficients computed for the electrode in the same position with all the others, averaged across trials.

Time-frequency analysis

For the bands of interest, the power was extracted every 10 ms within a 100 ms window for both the entire recording and the segmented data. Next, the mean spectrogram for each electrode was obtained by averaging across trials. The single-trial and mean spectrograms were then inspected in order to exclude any artefacts. The average spectrograms for beta and high gamma are reported in Figure 29 and Figure 30 for both MuSA and Epi devices, respectively. Both the high-gamma and the beta activation

spectrograms were were time-locked to the signal at the speech onset and segmenting the signals from 500 ms before to 500 ms after it. The spectrograms showed coherent activation between the MuSA and the Epi recordings. Furthermore, the time-frequency results confirm that µECoG recordings of neighbouring electrodes provide diverse information about the dynamics of neural activations during speech production: for the Epi recording, the power associated with the high-gamma range followed an apparent spatial activation pattern propagating from the bottom-right portion of the array, whereas the pattern associated with beta band spread from the top-left side of the matrix (see Figure 30). For the MuSA recording, the power extracted from the same bands shows different patterns but is only due to different orientations of the device on the cortex (see Figure 29).

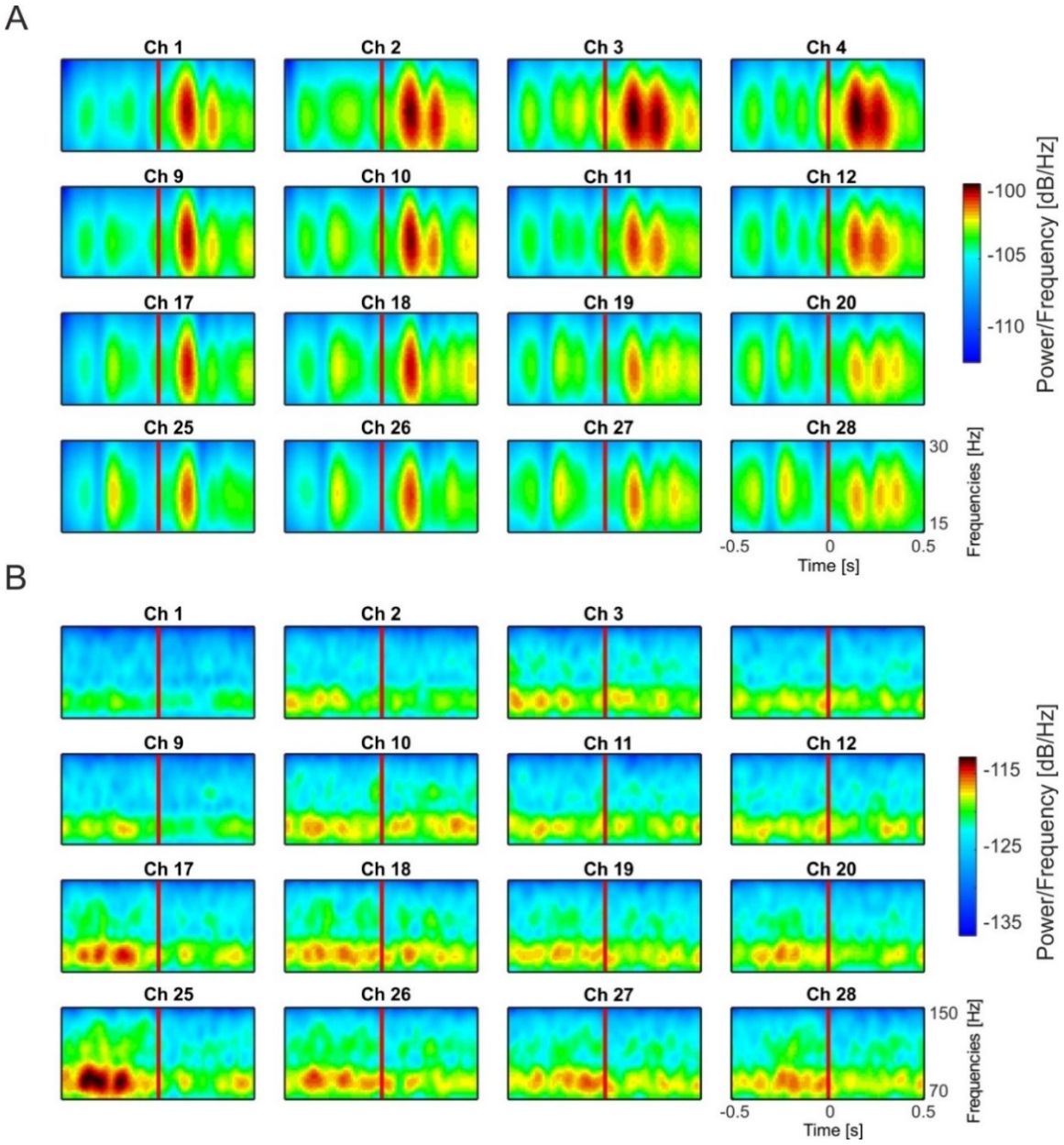


Figure 29. Averaged spectrogram maps for large electrodes of the MuSA array. Spectrograms were averaged across trials for the 16 electrodes in the range of frequencies 15-30 Hz (A) and 70-150 Hz (B). The window of analysis starts 500 ms before the onset of the speech and ends 500 ms after. The vertical red line represents the speech onset.

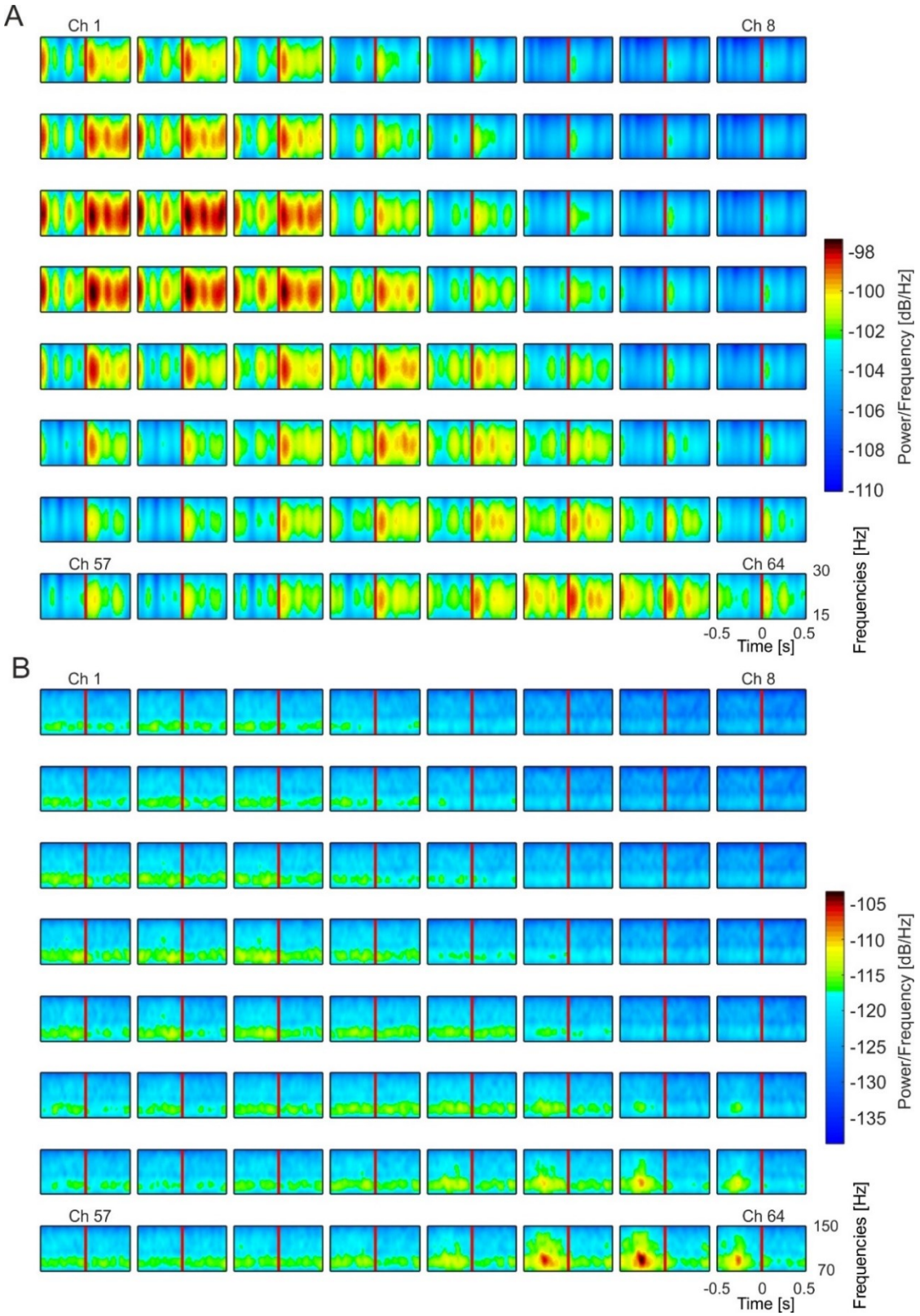


Figure 30. Averaged spectrogram maps for the Epi array. Spectrograms were averaged across trials for the 64 electrodes in the range of frequencies 15-30 Hz (A) and 70-150 Hz (B). The window of analysis starts 500 ms before the onset of the speech and ends 500 ms after. The vertical red line represents the speech onset.

Mean power profile

The mean power profile (MPP) was computed for high-gamma activity starting from the single-trial spectrograms to obtain a time-varying profile of the average activity in this frequency band. The results for the Epi and the MuSA devices are reported in Figure 31A-B, respectively, showing the most active channels selected on the basis of the mean spectrograms results.

Visual inspection of the single-trial MPPs allows us to identify the main characteristics of the preparation activity, which is fast, short and highly localized to few electrodes in a precise time window. The consistency of these features is demonstrated by the mean spectrograms. This prior knowledge was used to implement a simple but efficient method for predicting speech preparation from the neuronal signals.

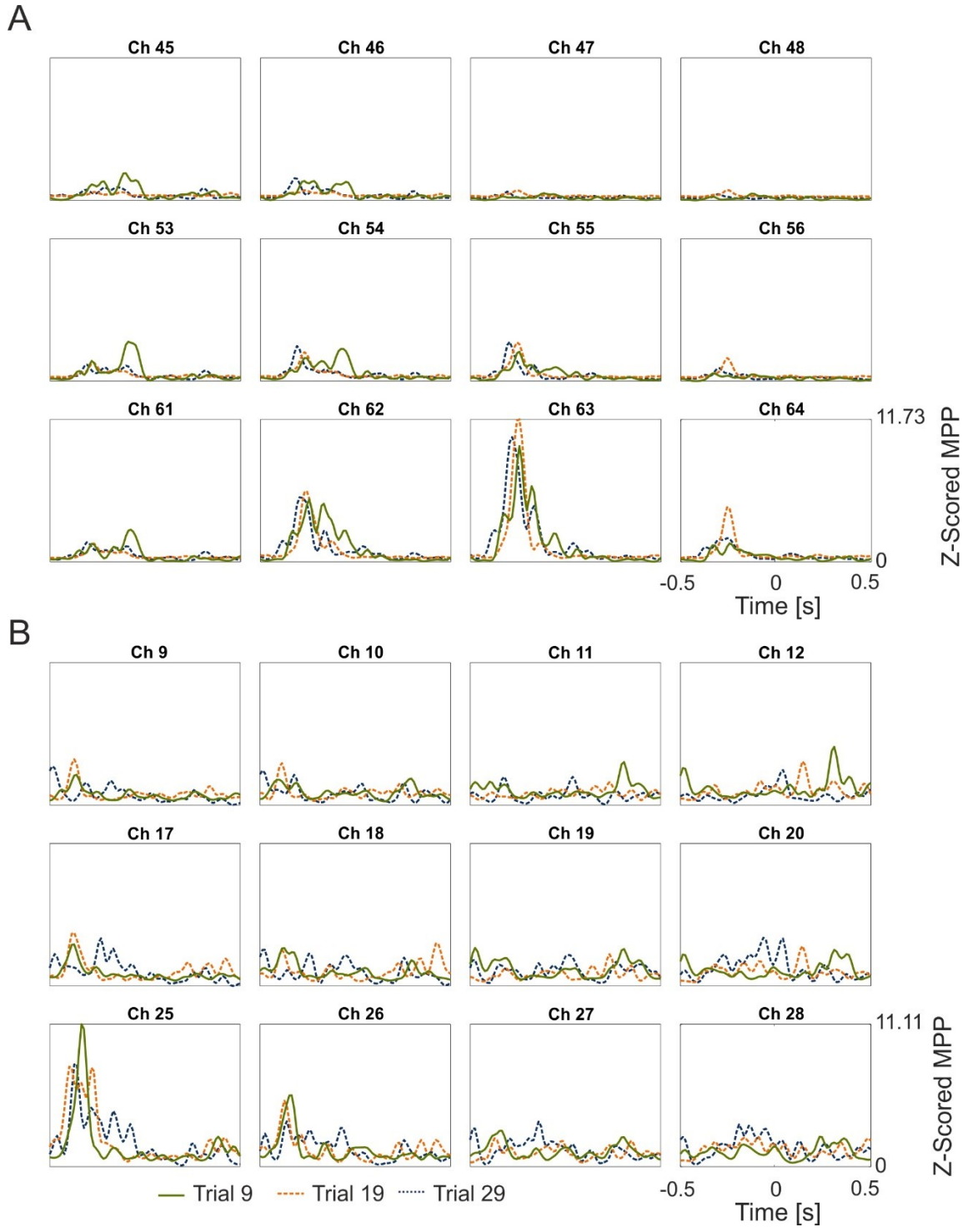


Figure 31. Mean power profile for some selected channels (see labels) of the Epi and the MuSA arrays. The spectrograms averaged across the frequencies in the range of 70-150 Hz are reported for the recordings performed with the Epi (A) and the MuSA (B). Each square represents a channel MPP during three different single-trials, represented with different colors and line-styles, aligned to the speech onset. The window of analysis starts 500 ms before the onset of the speech and ends 500 ms after.

3.4.1.2 Predicting speech preparation using SVM Classification

Speech preparation was predicted with a binary classification approach based on a support vector machine with gaussian kernel and leave-one interval-out cross validation. Hyperparameters were chosen from the recordings performed with the Epi, and the best combination was used to test the system with the MuSA recording. After extracting temporal features (MPP), I used 29 of 30 segments to train the classifier, leaving a different segment out at every iteration of the process. The trained classifier was tested on all the 60 ms shifting windows over the left-out interval, and all the test sets were then concatenated.

Dealing with unbalanced classes and context

As explained in sections *Feature extraction* and *Dealing with unbalanced classes*, the main issue in classification is balancing the number of observations with the number of features used for the training. In this case, given the low amount of data and the unbalanced classes, the choice of the two parameters, *window length* (i.e., number of features) and *proportionality factor (pf),* was made taking into consideration the Performance of Classification Index. The classification was performed by combining several values of *window length* and *pf*. The tests were conducted only for the Epi recording, since electrodes in this device showed more task-related activation.

The study of the different hyperparameters (*window length* and *pf*) allowed me to identify the best training conditions for the high-gamma features. Indeed, performance strictly depended on the *proportionality factor (pf)*, as well as on the number of features *(w)*; these consistently changed with the two hyperparameters (see Figure 32). The best performance obtained was 85.71 (*pf*=5, *window length*=3).

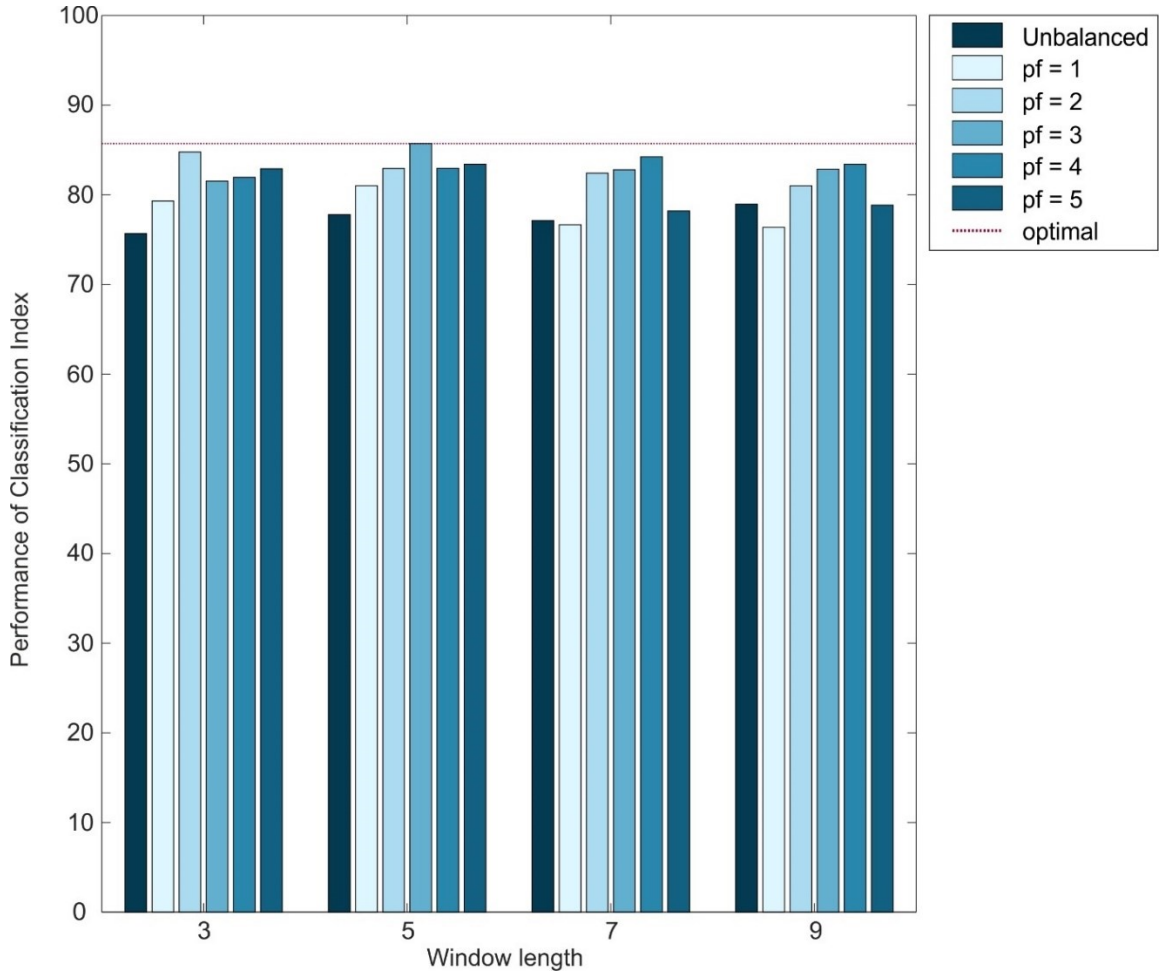


Figure 32. Performance of Classification Index for different proportionality factors and window lengths tested for high-gamma MPP features. The performance values obtained for high-gamma by varying the window length (x-axis) and the maximum imbalance between the classes (color-coded) are reported for n=1, taking into consideration only the best channel for each condition.

Classifier performances

It is well known from the literature that the brain activity recorded during cognitive tasks is characterized by an increase in the high-gamma band [3, 5, 16-18, 20, 97]. Evidence from this study demonstrates that speech production is also characterized by gamma and beta activation. To rule out the possibility that the other frequency bands might also allow prediction of speech preparation, the system was additionally tested with features extracted from beta and gamma bands, using the same hyperparameters used for high-gamma features. The results are reported in Figure 33. For each frequency band, i.e. beta, gamma and high-gamma, the PCI is reported for different tested *n* [1, 4]. In every subplot,

each single channel is represented as a square of the matrix following the spatial distribution of the electrodes in the device. The best-channel PCI reached by classifiers for n=1 was 16.08 for beta and 76.03 for gamma, compared to 85.71 reached with high-gamma features. These performance values are coherent with the diverse spatiotemporal patterns explored above with standard methods (spectrograms, correlations). This finding also demonstrates that the information extracted from adjacent electrodes may be profoundly variable in time and frequency at spatial resolutions below one millimeter.

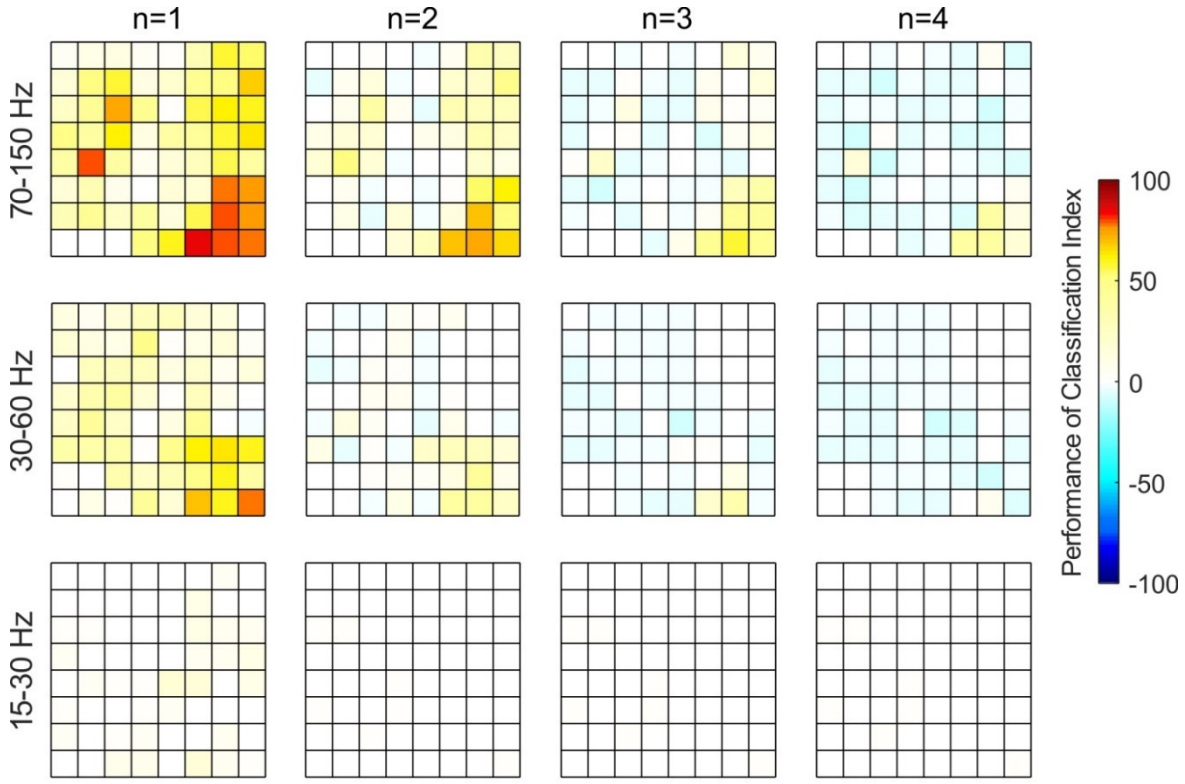


Figure 33. Prediction accuracy maps of the Epi array for MPP features. From the top, the performance obtained for high-gamma, gamma and beta are reported by varying n, the minimum number of samples that defines whether a trial was detected or not. The color scale represents the performance of classification index for each channel, represented as a square of the matrix.

The main complication of this study was evaluating our classifier performance from assignment of an ambiguous class such as speech preparation. The lack of information about the real timing of preparation intervals implies inevitable errors during the labelling procedure [68]. Nonetheless, I computed the standard confusion matrix and other derived

measures typically used to report classifier performance. The perfect classifier is expected to have an overall accuracy of 100%, meaning that it is both precise and sensitive. The *F-score* and *Matthews Correlation Coefficient* reach their best value at 1 (perfect classifier) and the worst at 0. For completeness, the classifier was trained also with randomized labels for the best combination of *window* and *pf,* and features were extracted from the high-gamma MPP.

The results of the different metrics and features are reported in Table 2 for the Epi and the MuSA recordings.

Table 2. Classification performance obtained for high-gamma features and optimal hyperparameters (window length and proportionality factor) with the Epi and the MuSA recordings. Only the values for the channel with the best PCI are reported.

	Epi		MuSA	
	True	Random	True	Random
Overall Accuracy	81.3%	79.3%	80.4%	80.5%
Matthews Correlation Coefficient	0.36	0.08	0.25	0.04
F-score	0.46	0.06	0.33	0.01
PCI	85.71	24.04	77.54	6.4

To visualize the predicted profile obtained from predicting for the best configuration (channel 62, *pf*=3, *window length*=5, F-score=0.46), the segments predicted during each test were concatenated and plotted with the true profile and the voice signal (see Figure 34). Interestingly, the classifier predicted also a segment during which the patient was reading a sentence from the screen, thus not part of the naming task (see the dark cyan segment).

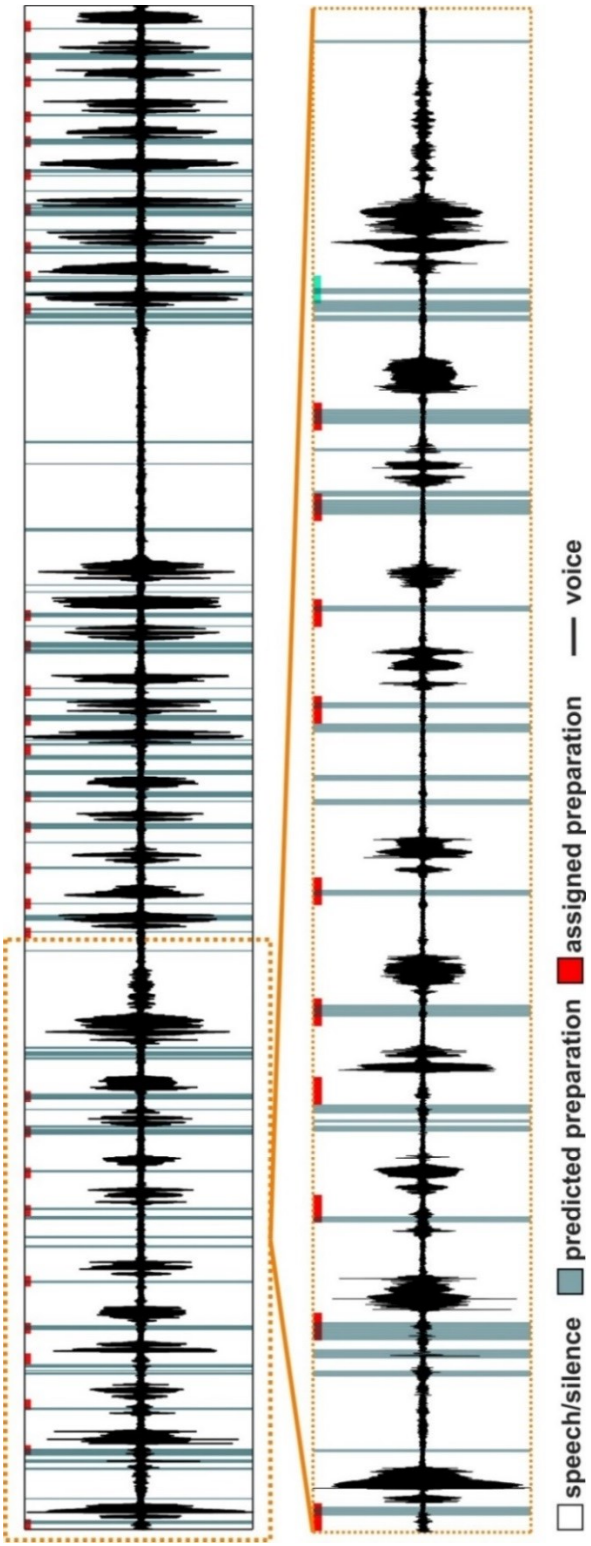


Figure 34. Predicted and hypothetical speech preparation profiles aligned with the voice signal. From the top, predicted labels for channel 62 and ideal labelling, aligned with the voice signal reported in black, are reported for the test sets (grey for predicted preparation, red for the real preparation and white for speech or silence). The bottom figure shows a zoomed interval, where the cyan interval underlines the presence of a speech preparation segment predicted by the classifiers even if not part of the naming task.

3.4.2 Discussion

In this case study, the recording micro-ECoG TDT-based setup was validated in a human subject undergoing awake neurosurgery. The experiment consisted of recording brain activity of the patient during performance of a naming task. The main goal was to demonstrate that the brain activity recorded with micro-ECoG devices, less invasive than standard ECoG grids, still carries relevant information and is diverse across electrodes for frequency and timing components.

Evidence from this study indicates that the neural signals are characterized by different temporal patterns whose spatial distributions depend on the frequency range of interest. According to previous literature, two critical aspects should be considered when studying correlation as a metric to define the resolution of specific phenomena:

1. correlations between pairs of electrodes are inversely proportional to their distance [77, 91];
2. correlations at low frequencies are greater than those at high frequencies, which are mainly recruited during the execution of tasks [97], such as speech production in Broca's area [5].

Since fast oscillations propagate around the signal source across few millimeters [98], high-density µECoG arrays should provide a more detailed picture of the phenomenon, improving the accuracy of neurophysiological processes description. These results provide evidence that broadens our knowledge about speech production processes in a language-related area at sub-millimeter scale. µECoG recordings of neighboring electrodes provide diverse information about the dynamic of neural activity during speech production. In addition, beta activity showed late activation aligned with the speech onset and was localized to several electrodes (see Figure 28B, Figure 29A, Figure 30A). In contrast, high-gamma brainwaves are characterized by a fast increase in power, involving few electrodes and anticipating the speech onset (see Figure 28D, Figure 29B, Figure 30B). Starting from the information provided by the correlation and time-frequency analyses, it seems fitting to conclude that high-gamma activation is the most confined in both time and space, while the energy in the beta bands propagates in a higher number of electrodes. Thus, neural activity recorded by sub-millimeter-spaced electrodes seems to modulate not only

the frequency band but also with the pitch and the position of the electrodes over the cortex (see Figure 28).

The second finding here may demonstrate that Broca's area is a key node in cortical networks responsible for speech production, participating during preparation rather than during articulation [5]. The consistency of high-gamma activations was shown using a classification experiment. Briefly, the high-gamma features were used to predict whether the patient was preparing to speak by training a Gaussian SVM classifier. I used a Leave-one trial-Out approach, since the number of trials was exiguous. The SVM classifier was trained with 29 trials and tested with the remaining one. This procedure was repeated by training a new classifier for each Leave-one trial-Out set, until all the trials were tested once.

The lack of information about the real timing of preparation intervals implies inevitable errors during the labelling procedure. This label noise made it difficult to evaluate our classifier performance. The results obtained for both the devices recording confirmed that the standard confusion matrix and other derived measures typically used to report a classifier performance, such as the overall accuracy, might not be relevant. These results are reported in Table 2 for both the Epi and the MuSA recordings. The overall accuracy seems to indicate that the performance could not reach or exceed randomized results. However, the PCI score showed values far from the randomized performances, similarly to the *F-score* and *Matthews Correlation Coefficient* behaviors. This demonstrated that the PCI is a reliable measure to evaluate the accuracy of the speech prediction system. In addition, the results obtained for the MuSA recording provide helpful insights into the spatial electrodes arrangement, which may inform fabrication of new µECoG devices. Indeed, the best performance was lower than the one obtained for the Epi recording. This finding suggests that the MuSA device was covering a sub-optimal and too circumscribed area. Ideally, the dimension and pitch of the electrodes should be optimized in a new device, taking into consideration its real application.

The resulting performance was also highly dependent on the imbalance between the preparation and non-preparation classes, as well as on the number of features. Reducing the number of samples for the majority class during the training changed the performance

values consistently, even though the best PCI performance was never reached in the completely balanced classes condition. This effect was a possible consequence of the limited number of samples available for the training. Moreover, the number of features also has a strong impact on the performance. Interestingly, the best classifier was able to detect two preparation segments that did not belong to the naming task (Figure 34). This was additional proof that the model was not over-fit to the naming preparation.

The same routine was also repeated for the MPP features extracted from gamma and beta bands. Gamma activity showed spatiotemporal patterns similar to those elicited in the high-gamma range, explaining the lower but sufficient PCI performances during classification. In contrast, the beta activity is recruited mainly after the speech onset; consistent with its temporal dynamic, the performance confirmed that this band is not well suited for predicting speech preparation.

4. CONCLUSIONS AND FUTURE PERSPECTIVES

ECoG-based speech BCIs could effectively help patients with severe communication disorders [31]. A growing body of literature has demonstrated that speech synthesis from neuronal activity is possible through ECoG recordings, paving the way for long-term speech-BCI systems [55, 56]. ECoG approaches are useful in BCI applications, as well as in clinical procedures, because this technique is characterized by high spatiotemporal resolution and low susceptibility to noise—both prerequisites for speech neuroprosthetic devices. However, there are both technical and computational issues that need to be improved.

One main issue is the stability and tolerability of chronic ECoG implants. The current approach is still relatively invasive, and the grid placement is strictly enforced by protocols for treating epilepsy and tumors. However, recent scientific developments may justify implantation in clinical populations that could benefit from such procedures, including BCI applications. Therefore, in order to increase the long-term performance, the flexibility and dimensions of ECoG grids should be drastically improved [31]. Furthermore, invasive BCI performance strictly depends on the inflammatory response and, thus, on dimension and stiffness of the device [29, 42, 79, 80]. Placing high-density micro-grids directly over language-related areas would increase the specificity of the recorded signals; additionally, adapting the device to conform to the brain would substantially reduce the risk of damage

to the cortex [77, 78]. In this study, I explored neural signals recorded in rats in to test a new generation of devices known as micro-ECoG (µECoG). I focused signals collected during *in vivo* experiments in rats to validate a µECoG-based setup. Here, two highly conformable polyimide-based µECoG devices were validated. The analyses provided evidence that µECoG devices can record high frequency multi-unit signals that are minimally affected by the small sizes of the electrodes and with a spatial resolution that has not been attained in previous cortical recordings. The potential to record high-frequency components of neuronal signals (above 200 Hz) directly from the cortex with a small and ultra-conformable device could be helpful in developing different applications, from understanding physiological networks activity to BCI research [29, 42, 79, 80].

The second key contribution of my thesis extends beyond the technical implantation limits. Speech neuroprosthetics aim to improve the quality of life of people suffering from communications deficits due to a variety of neurological causes. In these patients, communication might be not possible due to severe paralysis, even if language cortices are intact [81]. In such conditions, the speech-BCI system must be able to decode speech directly from brain activity [50, 55, 56, 64, 65, 67, 68]; thus, in order to start the speech decoding accordingly to language preparation, a neuronal cue might be necessary. In this study, I investigated the possibility of recognizing speech preparation from µECoG signals. For the first time, µECoG grids were acutely implanted in a human patient to investigate speech production processes in a language-related region. Our results revealed that during speech production, high-gamma activity shows well-defined temporal dynamics, with a related power augmentation occurring a few hundred milliseconds before speech onset. The brain activity elicited in this band was used to successfully train a support-vector machine (SVM) classifier for predicting speech preparation. These findings provide further indications that µECoG recordings provide access to valuable information at a very high spatiotemporal resolution, which could have important implications for designing speech-BCI devices. Indeed, the approach introduced here could be employed to trigger the speech decoding in the case of covert speech (see Figure 35). The results from analysis of human recordings reported here should be validated in more subjects, followed by being embedded in a speech decoder to test its effectiveness and reliability in a real-life scenario.

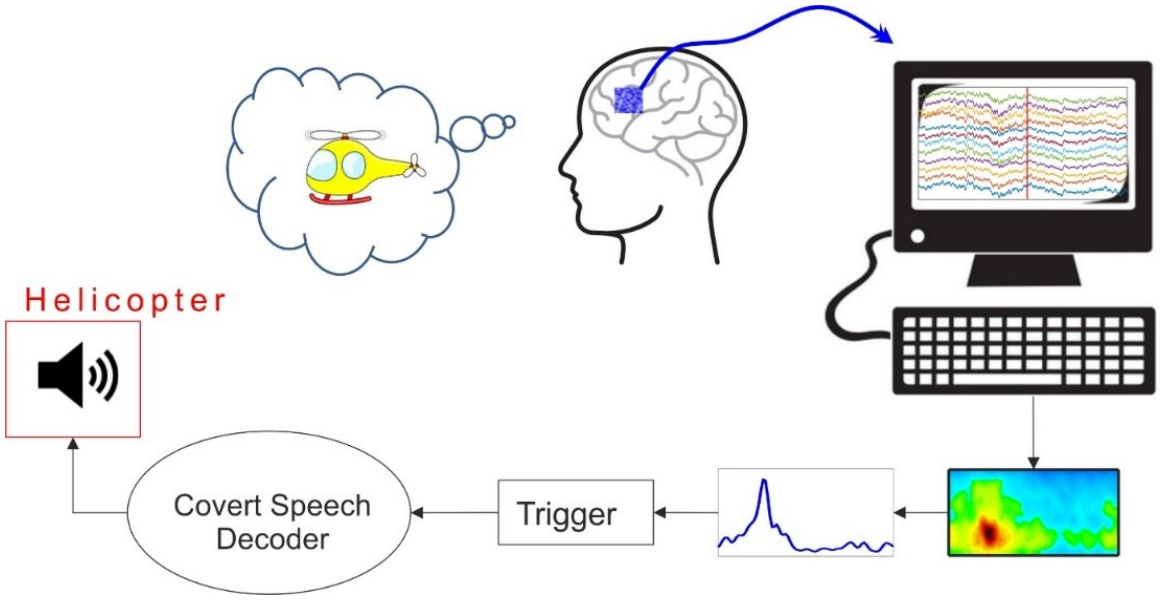


Figure 35. Overview of the ideal covert speech decoding system. The neural signals should be pre-processed online, and the speech prediction system would be employed as trigger for the covert speech decoder and voice synthesizer.

APPENDIX

A.1 Correlation analysis

The correlation analysis is a statistical method that allows quantifying the strength of the relationship between two continuous variables and it is usually applied to find a linear dependence between measures. If we consider two measures, *x* and *y,* both with *N* samples, the result of the correlation analysis is given by the *Pearson correlation coefficient*, which is defined as:

$$\rho(x,y) = \frac{1}{N-1}\sum_{k=1}^{N}\frac{x_k - \bar{x}}{\sigma_x}\cdot\frac{y_k - \bar{y}}{\sigma_y} = \frac{cov(x,y)}{\sigma_x\sigma_y}$$

By its definition, the Pearson correlation coefficient is the covariance of the two variables divided by the product of their standard deviations.

The correlation analysis has been previously used in Neuroscience to study the spatial distribution and propagation of neuronal activity in order to find the best compromise between cortex coverage and redundancy of information [77]. The choice depends on the frequency band of interest and, thus, also on the task. According to literature, two critical aspects need to be taken into account when studying functional correlation in neural networks related to language: 1) correlations between pairs of electrodes are inversely proportional to their distance [77] and 2) correlations at low frequencies are greater than at high frequencies, which are mainly recruited during the execution of tasks [97] like

speech production in Broca's area [5]. High-frequency components, which are believed to represent spatially localized neural processes, propagate around the signal source across few millimeters. Thus, the correlation coefficient is a key factor in the choice of the optimal ECoG electrodes spacing.

A.2 Fourier transform and time-frequency analysis in neuroscience

One way to explore signals in the frequency domain is the Fourier Transform, a mathematical formula that relates a signal sampled in time or space to the same signal sampled in frequency. In mathematics, Fourier analysis is the study of general functions approximation using sums of simpler trigonometric functions, sines and cosines, with three fundamental properties: frequency, power, and phase [99]. The Fourier transform produces a continuous function of frequency, known as a frequency distribution, starting from a time-varying signal. In signal processing, the Fourier transform can reveal important characteristics of a signal, namely, its frequency components. The Matlab function *spectrogram()* returns the short-time Fourier transform of the input signal *x,* providing a representation of the spectrum of frequency as it varies with time. The function is called using the parameters *window, noverlap, frequency_range*, *fs:*

$$[s, ps] = spectrogram(x, window, noverlap, frequency\_range, fs)$$

If *window* is an integer, then spectrogram divides x into segments of length *window*. Then, a Hamming window of that length is applied to each segment. *noverlap* is the number of overlapped samples, specified as a positive integer; if window is scalar, then *noverlap* must be smaller than *window*; *frequency_range* determines the frequency range of interest; *fs* is the sampling rate specified as the number of samples per unit time. The outputs are an estimate of the short-term, time-localized frequency content of x and an estimate of power spectral density (PSD) or power spectrum of each segment. Signal power as a function of frequency and time is a common metric used in signal processing for neuroscience [99]. In fact, EEG and ECoG data contain rhythmic activity [2, 98, 100] that can be analytically processed by means of amplitude, frequency and phase. These

properties are extracted from neural recordings using time-frequency analysis. Time-frequency analysis results have three dimensions and they can be represented as an image where time is on the x -axis and frequency is on the y –axis while the color of the plot (the z-axis) reflects some feature of the time-frequency data such as power.

A.3 Pattern classification

A.4 Dealing with unbalanced classes

Most real-world classification problems are characterized by imbalanced classes, i.e. the two classes are not equally represented in the dataset. The data imbalance is recognized as one of the major issues in the field of data mining. Indeed, most machine learning approaches assume that data are equally distributed; otherwise, the majority classes will dominate over minority classes causing classifiers to be more biased towards majority classes. This causes poor or zero detection of minority classes. Another consequence is the performance evaluation, because overall accuracy is not the best criterion to assess the classifier's performance in imbalanced domains. The goal to recognize speech-related motor preparation is an example of real-world classification problem suffering from imbalanced dataset. Possible approaches to deal with imbalanced classes consist in re-sizing the data by undersampling and oversampling [93] in order to change the prior conditions of the training set. Here, considering the small amount of data, overfitting was likely to occur and, therefore, I decided to adopt the *undersampling* approach, a popular technique for unbalanced datasets to reduce skew in class distributions.

A.5 SVM classification

The separation between speech preparation and not speech preparation segments was performed by *support-vector machines* (SVMs), which are widely described in literature and used when aiming to classify neural signals [53, 68, 70]. SVMs are a set of supervised learning methods typically used for classification problems when the observations cannot be linearly separable in their space.

The first step to create the SVM classifier is to choose a nonlinear function $\phi()$ (also known as *kernel function*) to map the observations *x* to a higher dimensional space in which the data points will be linearly separable by a hyperplane (see Figure 36A-B). The dimensionality of the mapped space can be arbitrarily high, but it is limited by computational resources. The resulting SVM classifier will provide a nonlinear decision boundary in the original space that maximizes the margin between the closest data points for the two classes known as *support vectors* (see Figure 36C). Other data points can be moved around freely (as long as they remain outside the margin region) without changing the decision boundary, and so the solution will be independent of such data points. The goal in training a Support Vector Machine is to find the separating hyperplane with the largest margin; we expect that the larger the margin is, the better the classifier generalization is.

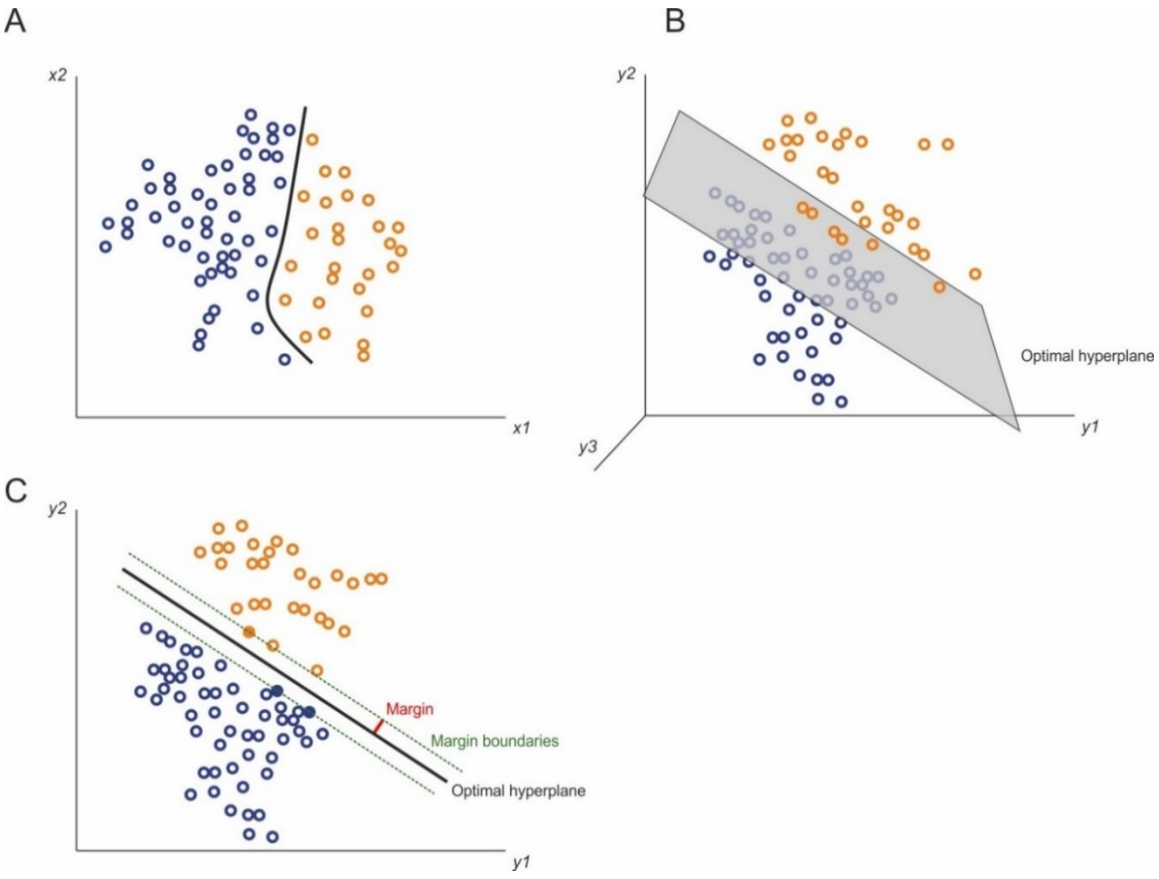


Figure 36. A) Example of nonlinear separation of two classes in a 2D space. B) Example of linear separation (hyperplane) of two classes in a higher dimensional space. C) The SVM training consists in finding the hyperplane with the maximum distance from the closest training data points, the support vectors. The support vectors are shown in solid dots.

A.6 Classifier performance

A binary classifier can be evaluated by means of standard metrics. The standard measures directly derive from the *confusion matrix,* which is represented as a matrix that summarizes the prediction results of a classification problem. As already explained, the confusion matrix can be extracted comparing the reference labels to the output of the classifier, i.e. the predicted labels. It is possible to define the standard metrics in terms of *true/false positive/negative.* The *true/false (T/F)* refers to the correctness of the predicted labels while *positive/negative (P/N)* refers to the label assigned by the classifier (i.e. class 1 and class 2). Using this terminology, it is possible to define the following parameters:

- *Sensitivity (True Positive Rate, TPR)* = $\frac{TP}{P}$
- *Specificity (True Negative Rate, TNR)* = $\frac{TN}{N}$
- *False Positive Rate (FPR)* = $\frac{FP}{N}$
- *False Negative Rate (FNR)* = $\frac{FN}{P}$
- *False Discovery Rate (FDR)* = $\frac{FP}{FP+TN}$
- *Precision (Positive Predictive Value, PPV)* = $\frac{TP}{TP+FP}$
- *F-Score* = $2 \cdot \frac{PPV \cdot TPR}{PPV + TPR}$
- *MCC*= $\frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)\cdot(TP+FN)\cdot(TN+FP)\cdot(TN+FN)}}$

REFERENCES

1. Kandel, E.R., et al., *Principles of neural science*. Vol. 4. 2000: McGraw-hill New York.
2. Buzsaki, G., *Rhythms of the Brain*. 2006: Oxford University Press.
3. Pei, X., et al., *Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition.* Neuroimage, 2011. **54**(4): p. 2960-2972.
4. Leuthardt, E., et al., *Temporal evolution of gamma activity in human cortex during an overt and covert word repetition task.* Frontiers in human neuroscience, 2012. **6**: p. 99.
5. Flinker, A., et al., *Redefining the role of Broca's area in speech.* Proceedings of the National Academy of Sciences, 2015. **112**(9): p. 2871-2875.
6. Petersen, C.C., *The functional organization of the barrel cortex.* Neuron, 2007. **56**(2): p. 339-355.
7. Knott, G.W., et al., *Formation of dendritic spines with GABAergic synapses induced by whisker stimulation in adult mice.* Neuron, 2002. **34**(2): p. 265-273.
8. Chen-Bee, C.H.-C., et al., *Whisker array functional representation in rat barrel cortex: transcendence of one-to-one topography and its underlying mechanism.* Frontiers in neural circuits, 2012. **6**: p. 93.
9. Hickok, G., J. Houde, and F. Rong, *Sensorimotor integration in speech processing: computational basis and neural organization.* Neuron, 2011. **69**(3): p. 407-422.
10. Di, S. and D.S. Barth, *Topographic analysis of field potentials in rat vibrissa/barrel cortex.* Brain research, 1991. **546**(1): p. 106-112.
11. Jones, M.S. and D.S. Barth, *Spatiotemporal organization of fast (> 200 Hz) electrical oscillations in rat Vibrissa/Barrel cortex.* Journal of neurophysiology, 1999. **82**(3): p. 1599-1609.
12. Kerr, J.N., et al., *Spatial organization of neuronal population responses in layer 2/3 of rat barrel cortex.* Journal of neuroscience, 2007. **27**(48): p. 13316-13328.
13. Petersen, C.C., *Sensorimotor processing in the rodent barrel cortex.* Nature Reviews Neuroscience, 2019: p. 1.

14. Castagnola, E., et al., *Smaller, softer, lower-impedance electrodes for human neuroprosthesis: a pragmatic approach.* Frontiers in neuroengineering, 2014. **7**: p. 8.
15. Brett-Green, B., et al., *Two distinct regions of secondary somatosensory cortex in the rat: topographical organization and multisensory responses.* Journal of neurophysiology, 2004. **91**(3): p. 1327-1336.
16. Crone, N.E., A. Sinai, and A. Korzeniewska, *High-frequency gamma oscillations and human brain mapping with electrocorticography.* Progress in brain research, 2006. **159**: p. 275-295.
17. Towle, V.L., et al., *ECoG gamma activity during a language task: differentiating expressive and receptive speech areas.* Brain, 2008. **131**(8): p. 2013-2027.
18. Crone, N.E., et al., *Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band.* Brain: a journal of neurology, 1998. **121**(12): p. 2301-2315.
19. Crone, N., et al., *Electrocorticographic gamma activity during word production in spoken and sign language.* Neurology, 2001. **57**(11): p. 2045-2053.
20. Crone, N.E., et al., *Induced electrocorticographic gamma activity during auditory perception.* Clinical neurophysiology, 2001. **112**(4): p. 565-582.
21. Sinai, A., et al., *Electrocorticographic high gamma activity versus electrical cortical stimulation mapping of naming.* Brain, 2005. **128**(7): p. 1556-1570.
22. Canolty, R.T., et al., *Spatiotemporal dynamics of word processing in the human brain.* Frontiers in neuroscience, 2007. **1**: p. 14.
23. Friederici, A.D., *The brain basis of language processing: from structure to function.* Physiological reviews, 2011. **91**(4): p. 1357-1392.
24. Broca, P., *Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole).* Bulletin et Memoires de la Societe anatomique de Paris, 1861. **6**: p. 330-357.
25. Tate, M.C., et al., *Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited.* Brain, 2014. **137**(10): p. 2773-2782.
26. Chang, E.F., et al., *Stereotactic probability and variability of speech arrest and anomia sites during stimulation mapping of the language dominant hemisphere.* Journal of neurosurgery, 2017. **126**(1): p. 114-121.
27. Mandonnet, E., S. Sarubbo, and H. Duffau, *Proposal of an optimized strategy for intraoperative testing of speech and language during awake mapping.* Neurosurgical review, 2017. **40**(1): p. 29-35.
28. Ferpozzi, V., et al., *Broca's area as a pre-articulatory phonetic encoder: Gating the motor program.* Frontiers in human neuroscience, 2018. **12**: p. 64.
29. Weltman, A., J. Yoo, and E. Meng, *Flexible, penetrating brain probes enabled by advances in polymer microfabrication.* Micromachines, 2016. **7**(10): p. 180.
30. Gunasekera, B., et al., *Intracortical recording interfaces: current challenges to chronic recording function.* ACS Chemical Neuroscience, 2015. **6**(1): p. 68-83.
31. Rabbani, Q., G. Milsap, and N.E. Crone, *The potential for a speech brain–computer interface using chronic electrocorticography.* Neurotherapeutics, 2019. **16**(1): p. 144-165.

32. Maynard, E.M., C.T. Nordhausen, and R.A. Normann, *The Utah intracortical electrode array: a recording structure for potential brain-computer interfaces.* Electroencephalography and clinical neurophysiology, 1997. **102**(3): p. 228-239.
33. Szostak, K.M., L. Grand, and T.G. Constandinou, *Neural interfaces for intracortical recording: Requirements, fabrication methods, and characteristics.* Frontiers in Neuroscience, 2017. **11**: p. 665.
34. Biran, R., D.C. Martin, and P.A. Tresco, *Neuronal cell loss accompanies the brain tissue response to chronically implanted silicon microelectrode arrays.* Experimental neurology, 2005. **195**(1): p. 115-126.
35. Hochberg, L.R., et al., *Neuronal ensemble control of prosthetic devices by a human with tetraplegia.* Nature, 2006. **442**(7099): p. 164.
36. Schwartz, A.B., et al., *Brain-controlled interfaces: movement restoration with neural prosthetics.* Neuron, 2006. **52**(1): p. 205-220.
37. Guggenmos, D.J., et al., *Restoration of function after brain damage using a neural prosthesis.* Proceedings of the National Academy of Sciences, 2013. **110**(52): p. 21177-21182.
38. Miller, K.J., et al., *Spectral changes in cortical surface potentials during motor movement.* Journal of Neuroscience, 2007. **27**(9): p. 2424-2432.
39. Chang, E.F., et al., *Categorical speech representation in human superior temporal gyrus.* Nature neuroscience, 2010. **13**(11): p. 1428.
40. Hill, N.J., et al., *Recording human electrocorticographic (ECoG) signals for neuroscientific research and real-time functional cortical mapping.* JoVE (Journal of Visualized Experiments), 2012(64): p. e3993.
41. Seymour, J.P., et al., *State-of-the-art MEMS and microsystem tools for brain research.* Microsystems & Nanoengineering, 2017. **3**: p. 16066.
42. Khodagholy, D., et al., *NeuroGrid: recording action potentials from the surface of the brain.* Nat Neurosci, 2015. **18**(2): p. 310-5.
43. Shenoy, P., et al., *Generalized features for electrocorticographic BCIs.* IEEE Transactions on Biomedical Engineering, 2007. **55**(1): p. 273-280.
44. Schalk, G. and E.C. Leuthardt, *Brain-computer interfaces using electrocorticographic signals.* IEEE reviews in biomedical engineering, 2011. **4**: p. 140-154.
45. Asano, E., et al., *Origin and propagation of epileptic spasms delineated on electrocorticography.* Epilepsia, 2005. **46**(7): p. 1086-1097.
46. Sugano, H., H. Shimizu, and S. Sunaga, *Efficacy of intraoperative electrocorticography for assessing seizure outcomes in intractable epilepsy patients with temporal-lobe-mass lesions.* Seizure, 2007. **16**(2): p. 120-127.
47. Miller, K.J., et al., *Real-time functional brain mapping using electrocorticography.* Neuroimage, 2007. **37**(2): p. 504-507.
48. Leuthardt, E.C., et al., *A brain–computer interface using electrocorticographic signals in humans.* Journal of neural engineering, 2004. **1**(2): p. 63.
49. Schalk, G., et al., *Two-dimensional movement control using electrocorticographic signals in humans.* Journal of neural engineering, 2008. **5**(1): p. 75.
50. Guenther, F.H., et al., *A wireless brain-machine interface for real-time speech synthesis.* PloS one, 2009. **4**(12): p. e8218.

51. Pei, X., et al., *Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans.* Journal of neural engineering, 2011. **8**(4): p. 046028.
52. Pasley, B.N., et al., *Reconstructing speech from human auditory cortex.* PLoS biology, 2012. **10**(1): p. e1001251.
53. Wang, Z., et al., *Decoding onset and direction of movements using electrocorticographic (ECoG) signals in humans.* Frontiers in neuroengineering, 2012. **5**: p. 15.
54. Herff, C., et al., *Brain-to-text: decoding spoken phrases from phone representations in the brain.* Frontiers in neuroscience, 2015. **9**: p. 217.
55. Anumanchipalli, G.K., J. Chartier, and E.F. Chang, *Speech synthesis from neural decoding of spoken sentences.* Nature, 2019. **568**(7753): p. 493-498.
56. Angrick, M., et al., *Speech synthesis from ECoG using densely connected 3D convolutional neural networks.* Journal of neural engineering, 2019. **16**(3): p. 036019.
57. Lebedev, M.A. and M.A. Nicolelis, *Brain–machine interfaces: past, present and future.* TRENDS in Neurosciences, 2006. **29**(9): p. 536-546.
58. Chaudhary, U., N. Birbaumer, and A. Ramos-Murguialday, *Brain–computer interfaces for communication and rehabilitation.* Nature Reviews Neurology, 2016. **12**(9): p. 513.
59. Ajiboye, A.B., et al., *Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration.* Lancet, 2017. **389**(10081): p. 1821-1830.
60. Wolpaw, J.R., et al., *Brain–computer interfaces for communication and control.* Clinical neurophysiology, 2002. **113**(6): p. 767-791.
61. Hochberg, L.R. and J.P. Donoghue, *Sensors for brain-computer interfaces.* IEEE Engineering in Medicine and Biology Magazine, 2006. **25**(5): p. 32-38.
62. Leuthardt, E.C., et al., *Using the electrocorticographic speech network to control a brain–computer interface in humans.* Journal of neural engineering, 2011. **8**(3): p. 036004.
63. Kao, J.C., et al., *Information systems opportunities in brain–machine interface decoders.* Proceedings of the IEEE, 2014. **102**(5): p. 666-682.
64. Moses, D.A., et al., *Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity.* Journal of neural engineering, 2016. **13**(5): p. 056004.
65. Martin, S., et al., *Decoding inner speech using electrocorticography: progress and challenges toward a speech prosthesis.* Frontiers in neuroscience, 2018. **12**: p. 422.
66. Brandman, D.M., S.S. Cash, and L.R. Hochberg, *human intracortical recording and neural decoding for brain–computer interfaces.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2017. **25**(10): p. 1687-1696.
67. Martin, S., et al., *Decoding spectrotemporal features of overt and covert speech from the human cortex.* Frontiers in neuroengineering, 2014. **7**: p. 14.
68. Martin, S., et al., *Word pair classification during imagined speech using direct brain recordings.* Scientific reports, 2016. **6**: p. 25803.
69. Wang, L., et al., *Analysis and classification of speech imagery EEG for BCI.* Biomedical signal processing and control, 2013. **8**(6): p. 901-908.

70. Wang, W., et al. *Decoding semantic information from human electrocorticographic (ECoG) signals*. in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2011. IEEE.
71. Akbari, H., et al., *Towards reconstructing intelligible speech from the human auditory cortex.* Scientific reports, 2019. **9**(1): p. 874.
72. Kellis, S., et al., *Decoding spoken words using local field potentials recorded from the cortical surface.* Journal of neural engineering, 2010. **7**(5): p. 056007.
73. Chen, Y., et al., *The 'when'and 'where'of semantic coding in the anterior temporal lobe: Temporal representational similarity analysis of electrocorticogram data.* Cortex, 2016. **79**: p. 1-13.
74. Mesgarani, N., et al., *Phonetic feature encoding in human superior temporal gyrus.* Science, 2014. **343**(6174): p. 1006-1010.
75. Leonard, M.K., et al., *Dynamic encoding of speech sequence probability in human temporal cortex.* Journal of Neuroscience, 2015. **35**(18): p. 7203-7214.
76. Yang, M., et al. *Speech reconstruction from human auditory cortex with deep neural networks*. in *Sixteenth Annual Conference of the International Speech Communication Association*. 2015.
77. Muller, L., et al., *Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography.* Journal of neural engineering, 2016. **13**(5): p. 056013.
78. Wang, P.T., et al., *Comparison of decoding resolution of standard and high-density electrocorticogram electrodes.* Journal of neural engineering, 2016. **13**(2): p. 026016.
79. Moshayedi, P., et al., *The relationship between glial cell mechanosensitivity and foreign body reactions in the central nervous system.* Biomaterials, 2014. **35**(13): p. 3919-3925.
80. Minev, I.R., et al., *Electronic dura mater for long-term multimodal neural interfaces.* Science, 2015. **347**(6218): p. 159-163.
81. Smith, E. and M. Delargy, *Locked-in syndrome.* Bmj, 2005. **330**(7488): p. 406-409.
82. Hickok, G. and D. Poeppel, *The cortical organization of speech processing.* Nature reviews neuroscience, 2007. **8**(5): p. 393.
83. Muller, L., et al. *Thin-film, high-density micro-electrocorticographic decoding of a human cortical gyrus*. in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2016. IEEE.
84. Castagnola, E., et al., *Biologically compatible neural interface to safely couple nanocoated electrodes to the surface of the brain.* ACS nano, 2013. **7**(5): p. 3887-3895.
85. Rembado, I., et al., *Independent component decomposition of human somatosensory evoked potentials recorded by micro-electrocorticography.* International journal of neural systems, 2017. **27**(04): p. 1650052.
86. Vomero, M., et al. *Achieving ultra-conformability with polyimide-based ECoG arrays*. in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2018. IEEE.
87. Vomero, M., et al., *Glassy Carbon Electrocorticography Electrodes on Ultra-Thin and Finger-Like Polyimide Substrate: Performance Evaluation Based on Different Electrode Diameters.* Materials, 2018. **11**(12): p. 2486.

88. Vomero, M., et al., *Highly stable glassy carbon interfaces for long-term neural stimulation and low-noise recording of brain activity.* Scientific reports, 2017. **7**: p. 40332.
89. Howell, D.C., *Statistical methods for psychology*. 2009: Cengage Learning.
90. Cohen, M.R. and A. Kohn, *Measuring and interpreting neuronal correlations.* Nature neuroscience, 2011. **14**(7): p. 811.
91. Rogers, N., et al., *Correlation structure in micro-ECoG recordings is described by spatially coherent components.* PLoS computational biology, 2019. **15**(2): p. e1006769.
92. Pastore, A., et al. *Motor Intention Decoding During Active and Robot-Assisted Reaching*. in *2018 7th IEEE International Conference on Biomedical Robotics and Biomechatronics (Biorob)*. 2018. IEEE.
93. Barandela, R., et al. *The imbalanced training sample problem: Under or over sampling?* in *Joint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR)*. 2004. Springer.
94. Angluin, D. and P. Laird, *Learning from noisy examples.* Machine Learning, 1988. **2**(4): p. 343-370.
95. Natarajan, N., et al. *Learning with noisy labels*. in *Advances in neural information processing systems*. 2013.
96. Bundy, D.T., et al., *Characterization of the effects of the human dura on macro-and micro-electrocorticographic recordings.* Journal of neural engineering, 2014. **11**(1): p. 016006.
97. Donner, T.H. and M. Siegel, *A framework for local cortical oscillation patterns.* Trends in cognitive sciences, 2011. **15**(5): p. 191-199.
98. Buzsáki, G., C.A. Anastassiou, and C. Koch, *The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes.* Nature reviews neuroscience, 2012. **13**(6): p. 407.
99. Cohen, M.X., *Analyzing neural time series data: theory and practice*. 2014: MIT press.
100. Buzsáki, G. and A. Draguhn, *Neuronal oscillations in cortical networks.* science, 2004. **304**(5679): p. 1926-1929.