

UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN BIOINGEGNERIA

Analysis of neural correlates of speech imagery for non-invasive brain-computer interfaces

Relatore

Prof. Luca Tonin

Laureanda

Margherita Gnocato

Correlatore

Tommaso Cortecchia

ANNO ACCADEMICO 2024-2025

Data di laurea 01/12/2025

Abstract

Language represents one of the most complex and, at the same time, fundamental cognitive functions in human life. However, certain neurological disorders can cause paralysis conditions that severely compromise communication with the external world, both verbal and gestural. To preserve quality of life under these conditions, various forms of Augmentative and Alternative Communication (AAC) have been developed. Nevertheless, currently available technologies present significant limitations, as they rely on slow and unintuitive paradigms, often unsuitable for the most severe cases of paralysis. In recent years, there has been growing interest in the development of speech brain-computer interfaces (speech BCIs), neural interfaces capable of decoding language directly from brain activity. The paradigm employed in such applications is speech imagery, i.e., the imagination of speech without overt sound production or articulatory movements. This approach offers greater intuitiveness and speed of use, making it a promising alternative to classical approaches based on event-related potentials or motor imagery. The present work focuses on the characterization of the neurophysiological correlates of speech imagery through the analysis of electroencephalographic (EEG) signals, with the ultimate goal of contributing to the development of a non-invasive speech BCI. After a brief overview of the main neurophysiological models of speech production and perception, as well as the current state of the art in speech BCIs, an experimental study is presented in which healthy participants performed overt and silent speech tasks involving sentences and isolated syllables. EEG signals were preprocessed using standard cleaning procedures, including spatial and frequency filtering, artifact removal through ASR and ICA, interpolation of noisy channels, and rejection of contaminated trials. Time-domain analysis of event-related potentials (ERPs) was then conducted. To reduce the risk of waveform cancelation due to phonetic variability, tasks were classified according to syllabic structure and articulatory features. The analyses revealed the presence of evoked potentials related to the experimental protocol. In contrast, in brain regions typically associated with linguistic functions, the observed activity does not appear to be interpretable as ERPs linked to syllabification during speech imagery. It is possible that the absence of strong evidence is due to the protocol employed and to the intrinsic complexity of the phenomenon, suggesting the need for future studies with more targeted experimental approaches.

Sommario

Il linguaggio rappresenta una delle funzioni cognitive più complesse e al tempo stesso fondamentali nella vita umana. Alcune patologie neurologiche possono però portare a condizioni di paralisi che compromettono la comunicazione con l'esterno, sia verbale che gestuale. Al fine di garantire una buona qualità di vita alle persone colpite da questa condizione sono state sviluppate forme di comunicazione alternativa (AAC). Le tecnologie ad oggi disponibili presentano tuttavia numerose limitazioni, in quanto si basano su paradigmi lenti, poco intuitivi e spesso non adatti ai casi più gravi di paralisi. Negli ultimi anni è emerso un crescente interesse per lo sviluppo di "speech brain computer interfaces" (speech BCI), interfacce neurali in grado di decodificare il linguaggio sfruttando i segnali neurali dell'utente. Il paradigma impiegato da questo tipo di applicazioni è lo "speech imagery", ovvero l'immaginazione del parlato senza produzione di suono o movimento. Tale approccio garantisce maggiore intuitività e velocità di utilizzo, rendendolo una potenziale alternativa agli approcci classici basati su potenziali evocati o "motor imagery". Il presente elaborato si concentra sulla caratterizzazione dei correlati neurofisiologici dello "speech imagery" attraverso l'analisi di segnali elettroencefalografici, con l'obiettivo finale di contribuire allo sviluppo di una speech BCI non invasiva. Dopo una breve rassegna dei principali modelli neurofisiologici della produzione e comprensione del parlato, e dello stato dell'arte delle speech BCI, viene presentato uno studio sperimentale in cui partecipanti sani hanno eseguito compiti di parlato in condizione "overt" e "silent" di frasi e sillabe isolate. I segnali EEG sono stati processati con procedure standard di pulizia, inclusi filtraggio spaziale e in frequenza, rimozione del rumore tramite ASR e ICA, interpolazione dei canali rumorosi ed eliminazione dei trial contaminati da artefatti residui. Successivamente è stata condotta un'analisi nel dominio temporale dei potenziali evento-correlati (ERP). Per ridurre il rischio di cancellazione dei segnali a causa della variabilità fonetica, i task sono stati classificati sulla base della struttura sillabica e delle caratteristiche articolatorie (consonanti occlusive, fricative, nasali; vocali alte, medie, basse). Le analisi hanno evidenziato la presenza di potenziali evocati legati al protocollo sperimentale. Nelle regioni tipicamente associate alle funzioni linguistiche, invece, l'attività osservata non appare interpretabile come ERP riconducibili alla sillabazione in speech imagery. È possibile che l'assenza di forti evidenze sia dovuta al protocollo adottato e alla com-

plexità intrinseca del fenomeno, indicando la necessità di studi futuri con approcci sperimentali più mirati.

Contents

1	Introduction	1
1.1	The Centrality of Language and Speech in human life	1
1.2	Systems of augmentative and alternative communication	2
1.3	Brain Computer Interfaces as alternative communication tools	3
1.4	The Potential of Speech Imagery BCIs	5
2	The neurophysiology of speech imagery	7
2.1	Linguistic models of speech production	7
2.2	From Speech Production to Speech Imagery	9
2.2.1	Cortex regions involved in speech imagery	9
2.2.2	Speech imagery in the frequency domain	11
2.2.3	Speech Imagery in the time domain	12
3	Existing Speech BCI	15
3.1	Invasive Speech BCI	15
3.2	EEG-based Speech Imagery BCI	16
4	Methods	19
4.1	Dataset	19
4.1.1	Experimental Setup	20
4.1.2	Experimental Design	20
4.1.3	Linguistic and Phonetic Characteristics of the Stimuli	21
4.2	ERP study	24
4.2.1	Preprocessing	24
4.2.2	ERP Grand Average analysis	25
4.2.3	ERP clustering	27
4.2.4	ERP classification	28
4.3	Time Frequency Feature Extraction and Classification	29

5	Results	31
5.1	Grand Average ERPs visualization	31
5.2	Subject level visualization	34
5.3	ERP clustering results	34
5.4	ERP classification results	36
6	Discussion and Limitations	39
6.1	Evoked Potentials	39
6.2	Speech Imagery ERPs	40
7	Conclusions	43
	Bibliography	45

List of Figures

4.1	Representation of the Experimental Protocol.	21
4.2	Distribution matrix of syllables across phonetic families. Each cell represents the frequency of syllables belonging to a specific phonetic class, defined by its manner (rows) and place (columns) of articulation. Marginal histograms indicate the cumulative frequency for each category.	22
4.3	Preprocessing pipeline.	25
5.1	Grand average ERP topoplots across syllable structures: V - vowels only; VC - syllables starting with vowels; CV - syllables starting with consonants.	32
5.2	Grand average ERP topoplots for syllables beginning with consonants, grouped by consonant phonetic families.	32
5.3	Grand average ERP topoplots for syllables beginning with vowels, grouped by vowels phonetic families.	33
5.4	ERPimage channel <i>AF3</i> , all trials. On the left plot, trials are sorted by subject; on the right, they are sorted according to the syllable code, which follows the sonority-based classification described in Subsection 4.2.2.	33
5.5	ERPimage channel <i>OZ</i> , all trials. On the left plot, trials are sorted by subject; on the right, they are sorted according to the syllable code, which follows the sonority-based classification described in Subsection 4.2.2.	34
5.6	Subject-level ERP topoplot. The channels showing differences in ERP shape between the three classes represented (V, VC and CV) are circled.	35
5.7	Contingency matrices in Grand Average obtained for the <i>FCZ</i> channel and using 9 clusters. The cells of the matrices show the frequencies with which syllables containing a phoneme from a certain family (y-axis) belong to the various clusters (x-axis). The frequencies are normalized by column to highlight the distribution pattern of phonetic categories within clusters.	35
5.8	Mean classification accuracy per EEG channel across subjects with standard deviation (gray bars). Red bars indicate the channel with above-chance average accuracies, while blue bars represent below-chance average performances.	36

5.9 (A) Scalp heatmaps represent the accuracy obtained per channel at the subject level. (B) Mean accuracy across subjects. 37

5.10 Scalp heatmap showing the count of channels with above-chance accuracy among all 10 subjects. 38

5.11 Example of correctly classified trials. 38

Chapter 1

Introduction

1.1 The Centrality of Language and Speech in human life

Communication is an essential function of human life. The spoken language, in particular, represents the primary tool through which we transmit information, intentions, and emotions rapidly and precisely. This unique capacity allows us to coordinate actions, negotiate, and establish strong social bonds. Moreover, language acquisition is central to cognitive development [1], [2], [3]. From birth, the human brain exhibits a specific activation for speech, indicating an innate predisposition to linguistic communication [4].

However, this fundamental function can be compromised by severe neurological conditions, such as stroke traumatic brain injury, or brainstem hemorrhage. These diseases can lead to a condition known as Locked-In Syndrome (LIS). In its classical form, LIS is characterized by quadriplegia, paralysis of the cranial nerves, anarthria, preserved consciousness, and vertical eye movements. The extent of these impairments can vary depending on the nature of the lesion or the stage of recovery. Indeed, LIS is typically categorized into complete LIS (CLIS), in which paralysis is total, including eye movements, and incomplete LIS, in which some oculomotor functions beyond vertical gaze are preserved[5].

As anticipated, Locked-In Syndrome is typically associated with traumatic or vascular events that cause damage to the ventral pons, but also neurodegenerative diseases such as amyotrophic lateral sclerosis (ALS) can also result in a LIS-like condition[6]. In the present work, the term LIS patients will be used to refer to individuals affected by quadriplegia, anarthria, and a lack of voluntary oculomotor control, regardless of the specific underlying medical condition. ALS also leads to a progressive deterioration of respiratory function, which is the leading cause of death in these patients. Artificial ventilation systems, both invasive and non-invasive, have therefore become increasingly important in prolonging life expectancy in ALS patients. However, invasive ventilation techniques that require a tracheostomy can further exacerbate communicative

difficulties [6], [7].

Thus, due to both the medical condition itself and the side effects of treatment, patients with LIS, despite preserved consciousness and, in many cases, intact cognitive functions, often find themselves unable to communicate, either through writing or speech. This has a profound impact on quality of life: patients lose the ability to express their daily needs, actively participate in therapeutic decision-making, and, in some cases, communicate their end-of-life wishes. For this reason, the development of alternative communication strategies has become a pressing necessity.

1.2 Systems of augmentative and alternative communication

The use of augmentative and alternative communication (AAC) technologies has been shown not only to improve patients' quality of life but also to positively influence the course of the disease. Furthermore, as mechanical ventilation systems are becoming increasingly widespread in many countries, it is likely that more ALS patients will progress to the most advanced stages of the disease. Given the current absence of curative treatments, an increasing number of patients will have to live with CLIS and the associated communication impairments [8]. The most commonly implemented solutions for patients who retain minimal voluntary movements rely on residual eye or eyelid movements in response to yes/no questions, the use of pen and paper, alphabet boards, or the selection, via small movements of limbs or eyes, of letters proposed by a communication partner. More advanced technological solutions include devices with screens combined with mouse or joystick control, which can be operated through reliable movements of the hands, feet, tongue, or eyes. When residual voluntary movements are highly limited and essentially binary (i.e., binary switches characterized by a "default" and an "activated" state), communication can be achieved through scanning interfaces: letters, words, or icons are presented sequentially, and the desired element is selected thanks to an activation signal produced by the patient. A particularly effective solution in such cases is eye-tracking technology, which allows for relatively rapid communication through the selection of words, letters, or icons based on eye movements. However, even these more technologically advanced systems are often unusable in practice. Although it is true that oculomotor control is often relatively spared in the progression of ALS, a significant proportion of patients present oculomotor deficits from the early stages of the disease. In other cases, the technology is impractical due to severe eye fatigue or difficulties in head stabilization. By contrast, low-tech AAC solutions are simple to implement in a domestic context but may lead to fatigue and frustration, as they are slow in transmitting information and generally require the constant presence of a trained communication partner familiar with the adopted AAC strategy [5], [6].

1.3 Brain Computer Interfaces as alternative communication tools

As can be observed, the mentioned solutions are based on the presence of residual movements and therefore do not assist patients affected by CLIS. Brain-computer interfaces (BCI) represent a viable solution even in the most severe cases of the syndrome: by completely bypassing the muscular system, they allow direct use of neural signals to interpret the user's intentions and restore a communication channel [6]. Specifically, this is achieved through a combination of hardware and software capable of acquiring brain data, processing it, and extracting meaningful parameters (features), which are then classified to determine the user's intended action. This can be accomplished through various methodologies, as each element of the pipeline can be implemented following different approaches.

A first classification can be made based on the invasiveness of the acquisition system. Invasive systems, such as those based on the acquisition of the electrocorticographic signal (ECoG), require the implantation of electrode arrays directly on the cortical surface (subdural implants) or just above the meningeal layer (epidural implants). These two systems involve surgical procedures with craniotomies and durotomies characterized by critical issues related to both the surgical procedure and the short and long term presence of electrodes. Non-invasive systems are based on neural signals such as electroencephalography (EEG) and near-infrared functional spectroscopy (fNIRS). This category also includes magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI), which, however, due to their high cost and lack of portability, are virtually unusable in a domestic setting. BCIs also differ in terms of paradigms, i.e., the pattern of interaction between user and machine: which feature is exploited, which area of the brain generates it, and how this signal is evoked, extracted, and interpreted. The choice of the appropriate paradigm depends on the required BCI application. BCIs can be used for rehabilitation systems, alternative communication, or control of robotic limbs, motorized wheelchairs, or cursors on a monitor. Below are some examples of how different EEG-based BCI have been implemented to date to establish an alternative communication system for people with ALS or LIS [5].

One of the first applications of neural interfaces for communication was a BCI based on slow cortical potentials (SCPs). The high voltage values of these slow waves are associated with states of cortical inactivity, while the low values are associated with movement or other cortical activity. In these studies from 1999 [9] patients learned to adjust SCP activation levels to move a cursor vertically and select the desired letter on a screen. The writing speed in this case was 2 characters per minute. [9], [10]. Similarly, BCIs have been designed to take advantage of the activation levels of sensory-motor rhythms (SMRs) corresponding to alpha

and beta rhythms [10-20Hz] and located in the sensory-motor cortex. These rhythms decrease in amplitude or desynchronize in preparation for movement, actual movement, or kinesthetic imagination of movement, and synchronize when movement is completed or at rest. After only a few training sessions, patients were able to correctly control sensorimotor rhythms and select one of the four targets presented on a screen. Using this paradigm, it is possible to increase the degrees of freedom related to screen selection by exploiting the desynchronization or synchronization of the two sides of the body (usually the hands). When sensorimotor rhythms are elicited by kinesthetic imagery of movement and in the absence of movement, this is referred to as Motor Imagery BCIs (MI BCIs).[10], [11] The Donchin Speller, a P300-BCI, differs from these BCIs in terms of paradigm and performance. This interface is based on the spontaneous response of our cortex to stimuli perceived as rare and unusual. This technology consists of a 6x6 matrix of letters and symbols whose rows and columns are illuminated randomly. The user selects the character simply by focusing on it. When the chosen character lights up, it is perceived as a rare stimulus, and the user's central cortex generates the P300 potential that gives the interface its name. Following numerous trials, the system identifies the presence of a P300 potential, and the classification component will determine the character desired by the user from the intersection of relevant rows and columns, providing it as feedback. In 1988 [12], when this technology was introduced, typing performance reached 2.3 characters per minute. The same paradigm can be used in the creation of auditory or vibrotactile BCIs, which are more suitable for assisting people with CLIS with oculomotor deficits but less powerful as they only allow binary selection. A fundamental difference between this technology and the two described above lies in the level of cognitive engagement required: compared to SCP- and MI-based BCIs, the Donchin Speller generally demands less sustained mental effort during use.[10], [12], [13] SSVEP-BCIs (Somato-Sensory Visual Evoked Potentials BCI) are interfaces that rely on the interpretation of visual potentials evoked by flashing light signals. In particular, neural signals oscillate at a frequency close to that of the flashing visual stimulus presented. This allows the user to passively select the desired icon from those presented on a display simply by staring at it. By analyzing the main frequency of the brain waves, the classification system will then be able to identify the user's intention. [14] Another difference between the applications presented lies in the origin of the signal used for classification: SCP- and MI-based paradigms are endogenous, meaning that the subject activates them when they wish to, rather than when a specific stimulus is presented from outside (P300-BCI, SSVEP-BCI). Endogenous BCIs therefore offer greater independence to the patient who, at least ideally, can choose when to start and stop communicating or writing.

1.4 The Potential of Speech Imagery BCIs

For years, BCIs have proven to be effective methods for re-establishing a form of alternative communication for people affected by LIS, but they do have limitations. BCIs based on a visual paradigm (Donchin Speller and SSVEP BCIs), for example, may be unsuitable for patients with CLIS due to the greater demand for visual activity, which is incompatible with oculomotor deficits. Furthermore, not all of them allow the creation of sentences; some only allow a choice between two alternatives. Interfaces that allow sentences to be formed without predefined options have writing speeds that are very different from spoken communication. Some of these issues could be resolved by BCIs based on “imagined speech”, “silent speech” or “speech imagery” (SI). BCIs that exploit this paradigm interpret the signals evoked by this attempt to speak without producing any sound or movement. The user would have to imagine the sound of their speech or the sensation of doing so, or both. This would cause neural signals that overlap partly with those generated during speech and that can be exploited to classify phonemes, syllables, or words to produce whole sentences. This type of BCI would have several advantages over currently existing technologies. First of all, it is an endogenous paradigm and would therefore allow spontaneous use, regulated entirely by the subject’s will. Furthermore, communication speed is affected by the acquisition mode used (ECoG or EEG), the subject’s level of training, and signal quality. In an ideal future scenario, a speech BCI could enable speeds much closer to those of natural speech (approximately 160 words per minute). [15]. In contrast, applications such as the Donchin speller have now reached their performance limits, as speed restrictions are intrinsic to the paradigm itself. The absence of the need for feedback or visual activities makes this type of BCI suitable even for patients with CLIS and less tiring for the eyes than SSVEP-BCI and P300-BCI. Furthermore, each of the classic applications presented in the previous paragraph relies on non-intuitive and unnatural paradigms for communication. Whether based on evoked potentials, such as the P300 potential, or on voluntary modulation of brain waves, none of these applications requires activity similar to that performed during speech. Therefore, what happens when MI-BCIs are used for the rehabilitation of specific movements compromised by pathologies or conditions does not occur. In the rehabilitation of stroke patients, numerous movements can be improved through the combined use of MI-BCI and functional electrical stimulation (FES). This seems to be precisely related to the similarity between neural activations during real and imagined movement and the association of this activity with visual and sensory feedback of movement caused by FES stimulation [16]. Therefore, the potential similarity of neural activation between SI and speech, combined with auditory feedback from the generated sentence, would lay the foundations for the development of rehabilitation techniques aimed at speech recovery in post-stroke patients. Furthermore, an activity more similar to natural speech could offer advantages in learning to use BCI. Creating sentences simply by thinking about them is

a more intuitive approach than thinking about moving a hand to select a letter, and more intuitive approaches can be learned more quickly as they exploit skills similar to pre-existing ones. This paradigm is not without its limitations and critical issues. The neural signals generated are notoriously difficult to decode because they are characterized by high spatial and temporal variability (inter- and intra-subject), low signal-to-noise ratio (SNR), and lack of behavioral output. [17] Examples and studies on speech BCIs, both invasive and non-invasive, that have actually been implemented, along with their limitations and potential, are discussed in detail later in this manuscript 3.

Chapter 2

The neurophysiology of speech imagery

Before specifically addressing the neural correlates of silent speech, the BCI paradigm under investigation in this thesis, some theories on the neurophysiology of speech will be presented. The aim is to derive a physiological rationale capable of guiding the numerous analyses required to develop a non-invasive speech BCI. The following paragraphs are therefore intended to address a series of questions: which cortical areas are involved in speech production and language processing? Which of these are also recruited during silent speech? Which of these generate activity that can be reliably measured through EEG? One of the reasons that motivates this focus on speech production is the assumption that there exists a partial overlap between the neural substrates engaged in overt speech production and those recruited during covert speech, analogous to what has been observed in the motor domain with motor imagery. Indeed, fMRI studies have shown that the cortical areas involved in motor execution and those associated with motor imagery show a significant degree of overlap.[18] Therefore, the expectation is to observe similar overlaps and correspondences in neural correlates of speech production and covert speech.

2.1 Linguistic models of speech production

Speech production and its control represent some of the most complex activities performed by the human brain. The muscles of the articulatory tract are extremely numerous, functionally redundant, and characterized by complicated biomechanical properties and multiple degrees of freedom. This complexity is also reflected by the fact that almost one third of the Primary Motor Cortex is dedicated to the control of muscles involved in speech production[19]. Their coordination gives rise to acoustic outputs that are not linearly related to the specific configurations of the vocal tract components. This occurs under the continuous influence of higher-order linguistic processes, such as semantics, syntax, and language-specific rules of sentence construction, which further increase the complexity of speech.

This complexity requires layered models of speech production. The general architecture of the most influential and empirically validated models is hierarchical in nature, comprising different functional modules: higher-level linguistic modules governing meaning and syntactic structure, a planning stage, a motor control stage, and the effector stage (the vocal and articulatory tracts). These modules interact through continuous bidirectional communication, involving both top-down and bottom-up pathways. The control stage plays a fundamental role in mediating between high-level linguistic functions and the articulatory execution of speech. The numerous models of speech production proposed to date differ in the specific implementation of the control stage, but they share several core characteristics. In general, the controller receives from the planning module a plan that, depending on the model, may be motor or acoustic in nature, representing the speaker's intended utterance in terms of articulatory actions or an auditory target. The controller integrates this intended plan with sensory and auditory feedback, originating, respectively, from the caudoventral precentral gyrus and the superior temporal gyrus, and generates corrective motor commands to minimize the mismatch between actual and predicted output.[20]

The following section briefly describes one of the most influential control models currently used in the literature: the DIVA (Directions Into Velocities of Articulators). The first formulation of the DIVA model is attributed to Frank H. Guenther (1994).[21], [22] Its current version postulates that the planning stage generates an acoustic representation of the utterance the speaker intends to produce. Based on this representation, three different types of trajectories are formed at the level of the cerebellum and thalamus: a motor trajectory, an auditory trajectory, and a somatosensory trajectory. The auditory trajectory is continuously compared, in real time, with the actual auditory feedback, producing an error signal in the event of a mismatch. An analogous comparison occurs for the somatosensory trajectory. In contrast, the articulatory (motor) trajectory is used as input to an internal forward model of the vocal tract. The output of this forward model corresponds to the predicted articulatory configuration, which is then compared to the actual configuration to detect discrepancies. The error signals resulting from these comparisons are transformed into corrective motor commands, which are then combined into a single command stream, the output of the whole model, that is relayed to the articulators of the vocal tract[20], [23]. This model has been able to account for behavior under both auditory and mechanical perturbations, to capture the phenomenon of motor equivalence in which multiple articulatory configurations can produce the same phoneme, and to remain consistent with known processes of speech acquisition[20].

To conclude this overview of speech production, the brain regions highlighted by the two main models will be briefly summarized. Understanding the structures involved is essential for generating physiologically grounded expectations regarding the regions of interest for the anal-

yses. Although there are minor differences in the specific regions recruited and their assigned functions, the core structures consistently identified include the motor and premotor cortex, the cerebellum, the Sylvian–parietotemporal area (Spt), and the superior temporal gyrus (STG).[22], [24]

2.2 From Speech Production to Speech Imagery

Similarly to the model described above, Tian and Poeppel (2010)[25] have also proposed a speech production model. In this framework, the motor plan is sent both to the articulatory effectors and, via an efference copy, to a first forward model. This model estimates the predicted somatosensory consequences of the planned action, which are then compared with the actual somatosensory feedback. At the same time, these predictions serve as input to a second forward model, which generates the expected auditory consequences of the motor commands. These auditory predictions are subsequently compared with the actual acoustic output, allowing on-line correction of motor commands. Building on this framework, Tian and Poeppel (2012)[26] further proposed a model describing what occurs during speech imagery. In this case, the simulation component implemented by the forward models remains intact; however, because no overt articulatory action is executed, the comparison with actual sensory feedback is absent. The auditory and kinesthetic feeling during the imagery task is therefore the consequence of this residual brain activity. These observations could suggest an involvement of the speech perception network during mental imagery of speech. This intuition has been supported by several studies employing different methodologies. For example, a recent fMRI study (2023) reported a substantial overlap between the brain regions significantly activated when participants listened to a poem being read aloud and when they imagined someone reading it.[27] Clearly, imagining someone else reading differs from imagining speaking oneself (i.e., covert speech). In covert speech, as discussed above, one would also expect the somatosensory experience of articulatory movements of the phonetic tract. Nevertheless, these findings are still promising because they suggest that, even in cases where participants may engage more strongly in auditory rather than somatosensory imagery, the speech perception system can still provide informative neural signals about covert speech.

2.2.1 Cortex regions involved in speech imagery

Tian and Poeppel (2013) [28] discuss a similarity between the auditory activity evoked by speech perception and the efference copies generated during speech production (SP) or speech imagery (SI) tasks. There are therefore points of contact between SI activity and SP activity: the areas involved and the neural signals generated in the two cases should be similar. From this per-

spective, models of SP become especially important. In particular, the most widely accepted framework for describing this organization is the Dual Stream Model (DSM) proposed by G. Hickok and D. Poeppel (2007) [29]. This model posits the existence of two processing streams: a ventral stream and a dorsal stream. The ventral stream mainly maps the perceived sound to its meaning, while the dorsal stream maps the sounds of speech into articulatory representations for speech production. As previously mentioned, these networks might also play a role in generating predictions during both overt and covert speech production. To describe how this happens, Tian and Poeppel (2013) [28] theorized the Dual Stream Prediction Model (DSPM), which involves the same networks described in the DSM for processing but with the information flow reversed: the dorsal stream maps the articulatory plan into phonological code, while the ventral stream maps from meaning into sound representations. Consequently, during SI tasks, the dorsal stream is recruited to generate the somatosensory feeling of the planned but not executed action, whereas the ventral stream is engaged to retrieve from memory the sound of the intended words. Summarizing the speech models discussed so far: as in voluntary motor control, speech production and control have been theorized to rely on internal forward models. These models involve neural networks that generate predictions of the expected auditory and somatosensory consequences of planned actions. Such predictions allow for real-time comparison between expected and actual sensory outcomes, supporting potential online correction. Although different models propose distinct mechanisms, they all share this core principle of simulation and anticipation. In SI tasks, the actual motor execution is suppressed, but the predictive components remain active. These predictive neural traces represent the primary target for speech BCIs, as they constitute a reliable neural surrogate for speech. The networks engaged in these predictive processes mirror those involved in SP described by the Dual Stream Model, but operate in the reverse direction. Therefore, the brain regions expected to be active during covert speech tasks are those described in the DSPM [28] and listed below:

- Inferior Frontal Gyrus (IFG, including Broca's area) and other premotor structures: involved in articulatory planning; The plan is normally sent to the motor cortex for execution, while the efference copy is forwarded to the Inferior Parietal Cortex. During covert speech the motor cortex is not recruited and only the efference copy is generated.
- Inferior Parietal Cortex: Responsible for estimating somatosensory consequences of speech movements in both overt and covert speech.
- Superior Temporal Gyrus (STG, including Wernicke's area): Encompasses the primary and secondary auditory cortices, where auditory predictions are represented in a spectrotemporal form and mapped onto phonological representations.

- Superior Temporal Sulcus (STS): supports higher-order auditory processing and the transformation toward phonological representations.
- Middle Temporal Lobe (MTL) and Middle Temporal Gyrus (MTG): part of the ventral streams that allows for memory retrieval and lexical access, supporting the auditory and semantic representations of internally generated speech.

The dorsal stream (which includes Broca's area, the premotor cortex, and the inferior parietal cortex) appears to be strongly left-lateralized, whereas the ventral stream (involving Wernicke's area, the STS, and temporal structures) tends to show more bilateral activation, with hemisphere-dependent differences in activation strength. To conclude this section, from a more practical and analysis-oriented perspective, we expect higher relevance in electrodes positioned over temporal, parietal, and frontal regions, with a possible left-hemispheric dominance. Overall, this section highlights that the neural substrates underlying overt and covert speech are largely shared, suggesting that the informational content of speech is preserved in a content-specific and temporally precise manner even in the absence of actual articulation or vocalization.[30] Moreover, considering how this neural activity is generated, some additional advantageous characteristics of speech imagery can be expected. In particular, as in overt speech, the individual manner of articulation is assumed to remain relatively stable over time. Furthermore, given the strong similarities between covert speech and speech perception processes, some authors[30] have hypothesized that speech BCIs could be trained using neural signals recorded during passive speech perception, rather than relying on more demanding covert speech tasks.

2.2.2 Speech imagery in the frequency domain

In terms of the regions of the cortex involved in SI activity, the relevant frequency bands could also be similar to the ones involved in speech processing. In SP, neural oscillations are involved in the theta, alpha, beta, and gamma frequency bands. Overall, the prevailing principle appears to be that slower oscillations (such as theta and alpha) are associated with the processing of longer linguistic units (phrases, words, and syllables) whereas faster oscillations, particularly in the beta and gamma ranges, are linked to the perception and comprehension of minimal speech elements, namely phonemes. It appears that, also in SI, gamma-band activity retains its correlation with the type of phonemes being produced, exhibiting a high degree of specificity for the imagined words.[19], [30] Another interesting aspect characterizing gamma-band activity during syllable production appears to be the phonotopic organization of the Superior Temporal Gyrus (STG). This cortical region exhibits spatial selectivity to the phonemes that are heard, reflecting a topographic mapping of phonemic features along its surface.[31] This property of the auditory cortex allows for perceived speech reconstruction through intracranial recordings.[32]

Poeppel and Assaneo (2020) [33] focus on the syllabic scale, an intermediate scale between the phonemic and word scales and shows how this mesoscale is fundamental to allowing speech comprehension and processing. In particular, it has been found that, largely independent of language, speakers, and speaking conditions, in natural speech syllables are produced at a frequency of approximately 2–8 Hz. Moreover, the alternation of consonants and vowels that characterizes syllables corresponds to moments of lower and higher sonority, respectively, which are clearly observable in the speech envelope that, consequently, also exhibits oscillations within this frequency range. Apparently, the listener’s neural activity in the theta band [4–8 Hz] becomes synchronized with the envelope of the perceived speech.[30], [34] In fact, when the syllabic rate is varied within the [2–8 Hz] range, a coherent shift can be observed in the peak of the theta-band spectrum, which synchronizes with the syllabic rhythm. Outside this range, synchronization with neural oscillations is lost, and speech becomes unintelligible.[35]. The study by Poeppel and Assaneo[33] also highlights a theta-band synchronization between auditory and motor areas. According to some interpretations, this motor rhythm corresponds to the rhythmic opening and closing of the jaw - with closures occurring at consonantal onsets and openings during vowel production. Other studies [30] have shown that, during SP, there exists a Phase–Amplitude Coupling (PAC) relationship between the theta and gamma bands: the phase of the theta-band oscillation modulates the amplitude of gamma oscillations. However, this characteristic appears to be absent during SI. The study identified the most informative frequency bands for discriminating between the perception and silent production of two words, “orange” and “blue”. During speech perception, several frequency bands were found to be significant, whereas in SI only higher frequency bands, such as beta and gamma, appeared to correlate with the imagined word. Furthermore, the theta–gamma PAC observed during speech perception was suppressed in SI, even showing an anticorrelation between theta phase and gamma amplitude. This difference seems to support the hypothesis that, in SP, theta-band activity reflects a motor–articulatory component (likely associated with rhythmic jaw movements) that is suppressed during SI. Interestingly, gamma-band activity in SI preserves a rhythmicity consistent with the theta oscillations observed during speech perception. This suggests that, during SI, gamma-band activity retains syllabic-level rhythmic modulation: its amplitude is still modulated by the phase of the envelope that would be generated during overt speech, despite the suppression of theta activity.

2.2.3 Speech Imagery in the time domain

The activity of listening and speech processing also manifests potentials that exhibit specificity for the phonemes heard.[31], [36], [37]. We are therefore talking about Phoneme Related Potential (PRP): ERPs with shapes and latencies that differ according to the combined phonemes. In particular, the most significant phonetic categorization seems to be based on the manner of

articulation [37]: how the sound is produced. For example, in plosive phonemes (in this dataset: /b/, /t/, /k/, /p/, /d/, /G/), the airflow is blocked at various points of articulation (such as the lips, alveolar ridge or velum) and then released with the production of sound. Vibrant consonants (/r/) are characterized by periodic and rapid alternation of such occlusions and releases of air. Similarly, other phonetic categories can be defined, such as nasals, affricates, fricatives, laterals, approximants, and different classes of vowels[38]. This topic will not be explored in depth in the present work. The several points of articulation of phonemes define the place of articulation, which, however, appears to be less relevant for distinguishing neural responses during speech perception.[37]

Given the similarity between SP and SI, a similar mapping between intended sound and ERP could also be seen in SI. However, in the context of SI, this characteristic has not been consistently reported in the literature. This may be more related to the difficulty of satisfying the assumption of time locking in SI activities. In fact, in the aforementioned studies[36], [37], the extraction of PRP waveforms is made possible by aligning epochs with acoustic features derived from the speech spectrogram, a procedure that is not possible in SI. Furthermore, in order to synchronize an SI task, it is necessary to use external stimuli that cause evoked activity (P300, SSVEP, Auditory Evoked Potentials) that can mask endogenous neural activity of interest. The extraction and analysis of ERPs associated with SI activity therefore represent a challenging and still unresolved problem.

Chapter 3

Existing Speech BCI

3.1 Invasive Speech BCI

Existing invasive BCIs are based on the use of electrocorticographic (ECoG) signals. These are obtained through arrays of microelectrodes placed directly on the cortical surface, in regions where neural activity is particularly discriminative for the intended application of the interface. In this section, only invasive BCIs that utilize speech imagery as a paradigm will be considered. In such cases, electrode arrays are typically positioned over the primary motor cortex and the ventral premotor cortex. The classification targets correspond to the residual neural activity associated with the planning of articulatory movements: individual movements related to facial or speech (even in the form of motor imagery) are mapped onto specific activation patterns of individual electrodes within the array [15], [39], [40]. Decoding these neural residues requires extremely high spatial resolution, as it is based on subtle differences in activation between adjacent microelectrodes, approximately 0.05 mm apart [15]. Due to this high spatial resolution and the superior information transfer rate, the performance of these interfaces is significantly higher than that of existing AAC systems. The current state-of-the-art interface [15] has achieved a communication rate of 62 words per minute, with an error rate of 23.8%, using a vocabulary of 125,000 words.

Clearly, these advantages are not without risks and complications. First, implantation of these arrays requires an invasive neurosurgical procedure involving both craniotomy and durotomy. Extensive preliminary studies are necessary to accurately identify the cortical target area for electrode placement. In addition, insertion of the electrodes into neural tissue induces inflammation and the subsequent formation of a glial scar, which progressively electrically insulates the electrodes from the surrounding neurons, leading to a gradual degradation of the device's performance over time[41], [42].

3.2 EEG-based Speech Imagery BCI

These significant limitations cannot always be considered negligible. For this reason, it is necessary to focus on non-invasive technologies, which are not only safer but also considerably more affordable and accessible. However, given the intrinsic differences between ECoG and EEG, it is unrealistic to expect that the methods described above for invasive applications can be replicated directly. Neural activity recorded from an area covered by a 3.2 mm microelectrode array becomes spatially averaged and loses any movement-specific information when captured through EEG. Unlike ECoG-based applications, which require extensive preliminary information to guide electrode implantation, EEG studies can take advantage of the whole brain coverage provided by this technology, allowing the simultaneous monitoring of multiple functionally relevant regions. This feature may even represent an advantage considering the widespread cortical distribution of areas thought to be involved in speech imagery activity.

Unlike ECoG-based BCIs, current EEG-based speech BCIs lack the flexibility to generate novel sentences or words. Existing systems are typically limited to the classification of a small set of predefined stimuli that were already included in the training datasets. Until now, existing studies have been limited to relatively simple classification tasks, such as distinguishing phonemes [43], syllables such as *yes* and *no*, differentiating between vowels [44], or discriminating among a small set of words or short phrases that may be useful only for basic communication needs. The classification methods used include traditional machine learning and deep learning approaches. The most commonly employed machine learning methods are Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA), applied to both binary and multiclass classification tasks. Other approaches include Random Forest, k-Nearest Neighbors (KNN), Common Spatial Patterns (CSP), Relevance Vector Machine (RVM), and Extreme Learning Machine (ELM). These methods can achieve good accuracy, but are generally applied to relatively simple classification problems and on datasets designed primarily to characterize the paradigm rather than to develop a functional BCI. Deep learning approaches appear to yield more promising results than traditional techniques and require less feature engineering, although they depend on much larger datasets. Several reasons may explain why a non-invasive BCI based on the SI paradigm has not yet reached practical usability.[45], [46] Analyzing the current state of the art in light of the neurophysiological evidence presented in Chapter 2, one major inconsistency emerges. A recent review [45] shows that very few studies assign importance to the gamma band. This is likely due to the fact that EEG recordings are highly contaminated by myographic activity in this frequency range, which often leads researchers to consider them unreliable. Furthermore, this may also stem from the tendency to conceptualize SI as a form of motor imagery (MI) that involves the articulatory tract, thus directing greater attention to the alpha and beta bands (typically associated with MI of the upper and lower limbs) rather than

to those frequencies that more specifically characterize speech-related activity. It is true that articulatory MI can still manifest in the beta band, but this does not necessarily guarantee that it carries enough discriminative information to distinguish words, syllables, or phonemes. In contrast, studies that include gamma-band activity in their analyses report a significant improvement in classification accuracy, highlighting its relevance for speech decoding[30], [47].

Another major limitation concerns data availability and methodological redundancy. As pointed out by [45], many studies in this field are based on the same few publicly available datasets, leading to numerous works that differ mainly in preprocessing or classification techniques rather than in experimental design. In addition, these datasets are often limited in terms of linguistic diversity, stimulus type, and recording duration, severely restricting their ecological validity and real-world applicability. For example, one of the most commonly used resources in this field is the KARA ONE dataset [48]. It contains multimodal recordings (EEG, face tracking, and audio) acquired during overt and silent pronunciation of seven phonemic/syllabic cues (*iy, uw, piy, tiy, diy, m, n*) and four words (*pat, pot, knew, gnaw*). Although studies based on this database introduce novel methodologies for speech imagery classification, the tasks on which they are developed remain far from those required to achieve a truly practical and useful application.

The present work is part of a broader and more ambitious research effort that uses a newly acquired EEG dataset, described in detail in Section 4.1, designed to enable the development of a functional EEG-based speech BCI. The main objective of this study is to characterize the neural correlates of speech imagery in the time–frequency domain, to extract insights that may support the development of EEG-based BCI systems and to assess the potential limitations of the data set itself.

Chapter 4

Methods

4.1 Dataset

The complete data set comprises 23 healthy adult subjects (9 females), between 22 and 33 years old according to the most recent update (12 September 2025). All participants were native Italian speakers with normal or corrected-to-normal vision and no history of neurological or speech disorders, and none of them had any prior experience with the speech-imagery paradigm before taking part in the recordings. Some of these analyses were conducted on a subset of subjects selected from the initial dataset. This subset consisted of 10 participants (4 females), between 24 and 28 years old, chosen based on their performance in the *rest vs. speech imagery* classification task carried out in a previous study. The performance of the subjects in question is shown in Table 4.1.

Table 4.1: Top ten performers in the Speech Imagery vs. Rest classification. Reported values correspond to the average accuracy and F1 score across models.

Subject	Accuracy (AVG)	F1-score (AVG)
f9	0.79	0.79
j4	0.73	0.71
k2	0.70	0.70
i5	0.71	0.69
j9	0.69	0.67
j3	0.63	0.67
j7	0.68	0.66
i1	0.65	0.64
j6	0.64	0.63
c7	0.64	0.62

This selection provides a degree of confidence that the analyzed data correspond to genuine

speech imagery activity occurring within the designated time windows.

4.1.1 Experimental Setup

The EEG dataset was acquired at a sampling rate of 2048 Hz using an ANT Neuro *eego*TM amplifier with active shielding, together with an ANT Neuro *waveguard*TM cap equipped with 64 electrodes (63 EEG + 1 EOG) arranged according to the international 10–20 system configuration. This standardized placement scheme defines electrode positions based on the relative distances between anatomical landmarks on the scalp (nasion, inion, and preauricular points), ensuring consistent spatial coverage of the cortical regions between participants. The recording reference was placed on the electrode *CPz*, and *AFz* served as the ground. Signal amplification was performed using an ANT Neuro EEG amplifier integrated into the *eego*TM system.

4.1.2 Experimental Design

For each participant, six runs were recorded in a single session: three in the *silent* condition (corresponding to speech imagery activity) and three in the *overt* condition (speech). Regardless of the condition, the structure of each run remained identical. Each run consisted of 20 trials, each associated with one sentence in Italian, totaling 60 sentences per condition.

The events related to each trial, with their corresponding code and the visual stimulus presented to the subject are reported below and are represented in Figure 4.1.

- **1. Start of the trial:** the screen was blank.
- **786. Fixation:** a fixation cross appeared in the center of the screen.
- **>1400. Cue phase:** the cross disappeared and the sentence corresponding to the current trial was displayed on the screen while the participant simultaneously listened to it through a Text-To-Speech (TTS) system. The order of presentation of the sentences was randomized.
- **782. Sentence rehearsal:** the sentence disappears and the participant has to repeat aloud (in the overt condition) or internally (in the silent condition) the sentence previously presented.
- **780. Preparation phase:** a short interval before the syllabic production phase.
- **781. Syllabic production:** a sequence of events, each lasting 600 ms, corresponding to the production of the syllables composing the sentence. Each syllable contained exactly one vowel and was presented on the screen using phonetic notation.

- **897. End of the trial.**

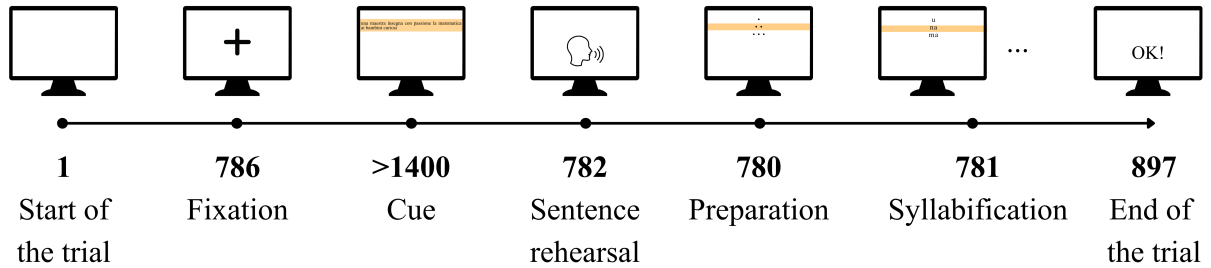


Figure 4.1: Representation of the Experimental Protocol.

Before the start of the recordings, participants were informed about the structure of the experimental protocol and about the use of phonetic notation for the presentation of syllables to avoid uncertainty during reading. Before data acquisition, participants were also shown the complete set of sentences they would later produce during the sessions. This familiarization phase lasted long enough for them to become comfortable with the stimuli but not so long as to allow memorization, which could have altered the cognitive processes associated with natural speech production. The first condition performed by each participant was the overt one, allowing them to become familiar with the selected sentences. Any errors occurring during the sentence rehearsal phase were noted and communicated to the participants between sessions to prevent them from being repeated in subsequent sessions and in silent condition, which was the most relevant for the purposes of this study. Before starting the silent condition, the participants were instructed on how to properly perform the speech imagery task: they were asked to imagine themselves on the verge of speaking, without actually articulating or vocalizing, while keeping their facial and articulatory muscles relaxed. It is important to note that the entire procedure lasted approximately two hours and that the speech imagery task is cognitively demanding. This inevitably led participants to experience fatigue during recording sessions.

4.1.3 Linguistic and Phonetic Characteristics of the Stimuli

The 20 sentences in the data set provide a representative coverage of the phonetic inventory of the Italian language [49], [50], while maintaining semantically neutral content to not evoke emotional reactions in the participants. Below is an example of one of the sentences included in the dataset, along with its English translation and syllabic segmentation. The whole sentence dataset used is reported in Table 4.2 at the end of this section.

Italian sentence: *una maestra insegna con passione la matematica ai bambini curiosi*

English translation: *a teacher passionately teaches mathematics to curious children*

Syllabic segmentation: *u.na ma.Es.tra in.seN.Na kon pas.sjo.ne la ma.te.ma.ti.ka a.i bam.bi.ni ku.rjo.si*

The sentences were chosen in order to ensure a balanced phonetic composition of the dataset, and syllabification was performed upon the phonemic transcriptions reported in *PhonItalia* [50]. Each syllable was categorized as shown in Figure 4.2. The figure shows the frequency distribution of syllables across the various phonetic families, classified according to their manner and place of articulation. The higher frequency of certain articulatory families reflects their natural prevalence in the Italian language [49], [50].

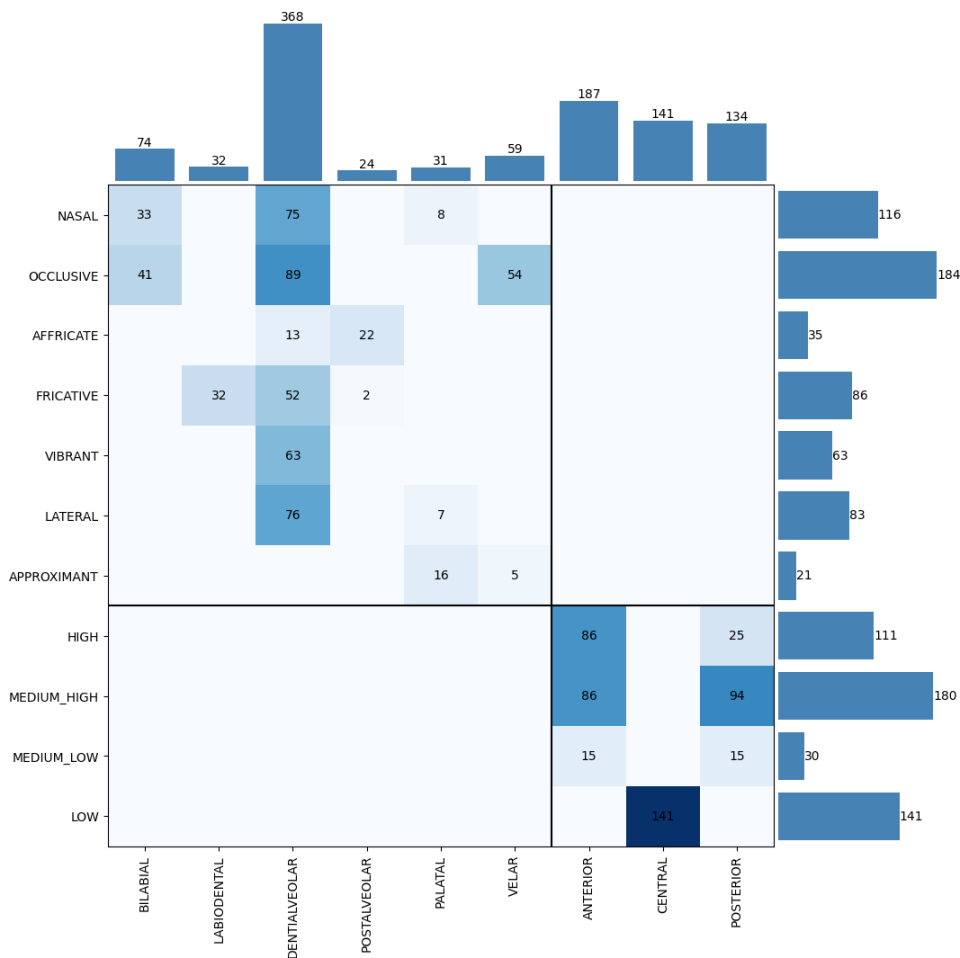


Figure 4.2: Distribution matrix of syllables across phonetic families. Each cell represents the frequency of syllables belonging to a specific phonetic class, defined by its manner (rows) and place (columns) of articulation. Marginal histograms indicate the cumulative frequency for each category.

Table 4.2: Sentences included in the dataset and their syllabic segmentation.

Sentences	Syllables boundaries
il bambino gioca felice con la palla rossa nel cortile assolato	il bam.bi.no gO.ka fe.li.ce kon la pal.la ros.sa nel kor.ti.le as.so.la.to
sua mamma sta preparando una torta dolce con cioccolato e fragole	su.a mam.ma sta pre.pa.ran.do u.na tor.ta dol.ce kon cok.ko.la.to e fra.Go.le
il papà guida lentamente la macchina blu sulla strada di montagna	il pa.pa Gwi.da len.ta.men.te la mak.ki.na blu sul.la stra.da di mon.taN.Na
la ragazza legge un libro interessante mentre beve il te caldo	la ra.Gaz.za leg.ge un li.bro in.te.res.san.te men.tre be.ve il te kal.do
il nonno racconta sempre storie divertenti ai nipoti prima di dormire	il nOn.no rak.kon.ta sEm.pre stO.rje di.ver.tEn.ti a.i ni.po.ti pri.ma di dor.mi.re
dei cani corrono nel prato verde dietro la grande casa bianca	de.i ka.ni kor.ro.no nel pra.to ver.de djE.tro la Gran.de ka.sa bjan.ka
il professore sta spiegando un argomento difficile ai suoi studenti	il pro.fes.so.re sta spje.Gan.do un ar.Go.men.to dif.fi.ci.le a.i swO.i stu.dEn.ti
il gatto osserva silenzioso gli uccellini sui rami degli alberi	il Gat.to os.sEr.va si.len.zjo.so Li uc.cel.li.ni su.i ra.mi deL.Li al.be.ri
ci sono ancora tante cose che non sappiamo sul cervello umano	ci so.no an.ko.ra tan.te kO.se ke non sap.pja.mo sul cer.vEl.lo u.ma.no
mio zio compra formaggio e latte freschi dal negozio qui vicino	mi.o zi.o kom.pra for.mag.go e lat.te fres.ki dal ne.GOz.zjo kwi vi.ci.no
mia zia cucina pasta al pomodoro con basilico e un po di olio	mi.a zi.a ku.ci.na pas.ta al po.mo.dO.ro kon ba.si.li.ko e un po di O.ljo
il treno ha lasciato la stazione centrale e arriva dopo un ora	il trE.no a laS.Sa.to la staz.zjo.ne cen.tra.le e ar.ri.va do.po un o.ra
il medico visita un paziente e gli prescrive una cura efficace	il mE.di.ko vi.si.ta un paz.zjEn.te e Li pres.kri.ve u.na ku.ra ef.fi.ka.ce
una maestra insegna con passione la matematica ai bambini curiosi	u.na ma.Es.tra in.seN.Na kon pas.sjo.ne la ma.te.ma.ti.ka a.i bam.bi.ni ku.rjo.si
il vento soffia forte tra gli alberi alti e muove le foglie	il vEn.to sof.fja fOr.te tra Li al.be.ri al.ti e mwO.ve le fOL.Le
il ragazzo ascolta la musica allegra mentre cammina verso la scuola	il ra.Gaz.zo as.kol.ta la mu.si.ka al.lE.Gra men.tre kam.mi.na vEr.so la skwO.la

Sentences	Syllables boundaries
la signora annaffia i fiori colorati nel giardino dietro la casa	la siN.No.ra an.naf.fja i fjo.ri ko.lo.ra.ti nel gar.di.no djE.tro la ka.sa
il bambino beve un succo dolce e mangia un panino al formaggio	il bam.bi.no be.ve un suk.ko dol.ce e man.ga un pa.ni.no al for.mag.go
un postino ha consegnato una lettera alla donna con il cappotto	un pos.ti.no a kon.seN.Na.to u.na lEt.te.ra al.la dOn.na kon il kap.pOt.to
la neve copre il tetto delle case e brilla sotto il sole freddo	la ne.ve kO.pre il tet.to del.le ka.se e bril.la sot.to il so.le fred.do

4.2 ERP study

Given that, in SI, the synchronization between theta-band activity and the rhythmic structure of speech (i.e., the periodic alternation of higher sonority phonemes - vowels - and lower sonority phonemes - consonants) appears to be absent, it would be valuable, for the purpose of building an EEG-based BCI, to identify a neural correlate that marks the timing of syllable production during SI. Such a marker would provide a time-locking cue to define short analysis windows in which to concentrate syllable classification. As mentioned in Section 2.2.3, from a perceptual point of view, the manner of articulation seems to be more relevant than the place of articulation in the neural encoding of phonemes. As a consequence, this part of the present work will focus on verifying if the ERP related to the syllabation in the silent condition shows some correlation with this categorization of syllables. To perform this part of the study only the subset of top ten subject mentioned in Table 4.1 in the SI vs rest classification were considered.

4.2.1 Preprocessing

The preprocessing pipeline followed a standard procedure and, as in the rest of the analyses, was performed in MATLAB (version 2024a) and using the EEGLAB toolbox[51]. First, the channels corresponding to mastoids (*M1* and *M2*) and EOG electrodes were removed, as they were noisy and not useful to improve signal quality in the remaining channels. Frontal and occipital electrodes (*FP1*, *FP2*, and *FPz*) were also discarded, as they were highly contaminated by ocular artifacts and were unlikely to provide relevant information for subsequent analyses. The three different runs acquired for each subject and condition (silent and overt) were concatenated and formatted to be compatible with the EEGLAB toolbox. To eliminate line noise artifacts, notch filters at 50 Hz and 100 Hz were applied, each with a normalized bandwidth of 35. In addition, a second-order Chebyshev band-pass filter with a passband of [1–20] Hz and a 1 dB passband

ripple was applied. Both filters were zero-phase Infinite Impulse Response (IIR) filters, used to avoid phase distortion. A first resampling step was then performed, reducing the original sampling frequency from 2048 Hz to 512 Hz. The EEG signal was then visually inspected in its entirety to identify channels exhibiting corrupted activity, likely due to errors in electrode placement or poor contact. These channels were subsequently interpolated using information from the neighboring electrodes. To perform a first cleaning of myographic contaminations, ASR was applied. To reduce the computational cost of the subsequent analysis, another down-sampling to 256 Hz was performed. To reject the artifacts, ICA was applied. The removal of the independent component was performed manually following a conservative approach with the objective of eliminating only those clearly associated with artifacts, in particular, activity in the eyes. The data were then divided into 600ms length epochs in correspondence of the events of syllabation (code 781). The signals still showing noise at this point were discarded. For each subject, the maximum amount of discarded epoch was 26 in a total of 1386 trails. To reduce the effects of volume conduction that are typical of EEG recordings and enhance spatial resolution, a laplacian spatial filter was applied.

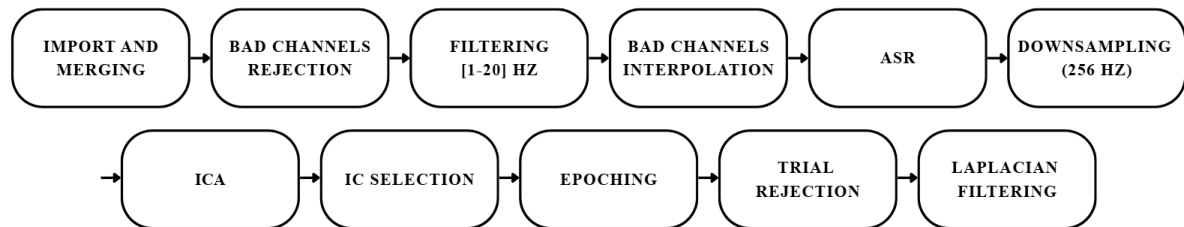


Figure 4.3: Preprocessing pipeline.

4.2.2 ERP Grand Average analysis

Data related to syllable events were categorized according to the phonemes that make up each syllable and their order. Each phoneme was first assigned to one of nine phonetic families, defined as follows:

Table 4.3: Phonetic family classification and corresponding phonemes. Each family is identified by a numerical code used in the syllable encoding scheme (**2XYZ**). The phonemes are reported as shown in the screen during the acquisition protocol

Code	Family name	Phonemes
1	Occlusive	/b/, /t/, /k/, /p/, /d/, /G/
2	Fricative and Affricate	/f/, /s/, /c/, /z/, /g/, /v/
3	Nasal	/m/, /n/, /N/
4	Lateral	/l/, /L/
5	Vibrant	/r/
6	Approximant	/j/, /w/
7	High vowel	/i/, /u/
8	Medium vowel	/o/, /O/, /e/, /E/
9	Low vowel	/a/

The numerical order of these families reflects their relative degree of sonority, ranging from the least sonorous (#1. Occlusive) to the most sonorous (#9. Low vowel). To obtain exactly nine categories, the Fricative and Affricate families were merged, as the phonemes of these groups share similar articulatory and sonority features, and the Affricate family exhibits low marginal frequency, as shown in Figure 4.2. The characters shown in Table 4.3 are exactly those used in the protocol and those used by *Phonitalia*,[50] the source from which the sentence dataset was extracted, which do not correspond exactly to those proposed by the International Phonetic Alphabet (IPA). Each syllable was then assigned a four-digit code (**2XYZ**), representing the phonetic composition of the syllable according to the families above:

- **X**: family of the first phoneme of the syllable
- **Y**: family of the second phoneme (0 if absent)
- **Z**: family of the third phoneme (0 if absent)

For example, the syllable *la* is assigned the code **2490**, while the syllable *djE* corresponds to **2168**, and so on. Accordingly to this categorization, the ERPs were calculated by averaging the z-scored trials of the various categories across all subjects. The Grand Average ERPs topoplots were then visualized comparing their structure: Vowel + Consonant (VC), Consonant + Vowel + * (CV) or only vowel (V). Further comparison was made in the CV category distinguishing between consonant families and in the V and VC categories distinguishing between high, medium, or low vowels. Taking into account the high variability between subjects that characterizes SI activity, the analysis was replicated at the subject level, and butterfly plots and ERP images per channel were displayed to establish whether the displayed average was representative of the trials. ERP images, in particular, were displayed according to the order of execution and

according to the order dictated by the coding system described in the previous paragraph 4.2.2. This display allows us to highlight whether the waveform is influenced by the passage of time (effect of fatigue or different runs) or by the type of syllables produced.

4.2.3 ERP clustering

To formally establish whether there was any correlation between the extracted waveforms and the syllabic families, similar to the study mentioned above [37], unsupervised hierarchical clustering was performed both on subject level data and on grand average data, considering all trials regardless of their syllabic type. The pairwise distances between the trials were calculated using the Euclidean metric explicitly defined in the following equation (other metrics, such as the correlation distance, were tested but produced comparable results).

$$d_{ij} = \sqrt{\sum_{t=1}^T (x_{it} - x_{jt})^2} \quad (4.1)$$

where x_{it} and x_{jt} are the ERP amplitudes of trials i and j at time sample t , and T is the total number of samples.

Clustering was performed using Ward's linkage method, an iterative algorithm which minimizes the total within-cluster variance at each step:

$$D_{(r,s)} = \sqrt{\frac{2n_r n_s}{n_r + n_s}} \|\bar{x}_r - \bar{x}_s\| \quad (4.2)$$

where n_r and n_s denote the number of elements in the clusters r and s , and \bar{x}_r and \bar{x}_s are their respective centroids. To explore the structure of the data, clustering was repeated independently for each EEG channel, systematically varying the number of clusters between ($nCl = 2$) and ($nCl = 18$). Rather than imposing a fixed number of clusters, this procedure was designed to avoid limiting the results with a priori assumptions about the possible relationship between phonetic families and ERP morphology. In principle, ERPs could capture both very subtle differences, such as those relating to individual phonemes, and broader patterns, such as syllabic structure, potentially leading to different optimal numbers of clusters. For each cluster configuration, three chi squared tests were performed: one for each of the three digits that make up the four-digit syllable codes described in Section 4.2.2. Each test assessed the null hypothesis that the clusters obtained were independent of the phonetic family corresponding to the position of that digit in the code. To assess whether the clustering captured meaningful phonetic information, the chi-squared test was applied, testing the null hypothesis that the clusters and the two classes were independent.

The test statistic was computed as:

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (4.3)$$

where O_{ij} is the observed frequency in cell (i, j) , and the expected frequency under independence is:

$$E_{ij} = \frac{(\sum_j O_{ij})(\sum_i O_{ij})}{N} \quad (4.4)$$

with $N = \sum_{i,j} O_{ij}$. The resulting p -value was obtained from the cumulative distribution function of the chi-square with $(R - 1)(C - 1)$ degrees of freedom and compared to the significance threshold of $\alpha = 0.05$. This data-driven approach was designed as a straightforward way to verify whether a strong distinction existed between specific phonetic categories of syllables, which could then be explored in more detail through further analysis.

4.2.4 ERP classification

To simplify the analysis and assess whether the ERP data contained information related to the syllables produced, trials corresponding to syllables with markedly different structures were selected: those beginning with an occlusive consonant followed by a vowel (hereafter referred to as OV syllables), versus those beginning with a vowel not followed by an occlusive consonant (hereafter referred to as V^* syllables).

In this analysis, a template matching classification algorithm was implemented to distinguish between two syllable classes based on the morphology of their ERPs. For each subject, the preprocessed data were divided into two classes which were then balanced by randomly subsampling the larger class. The data were subsequently z-scored to reduce the impact of amplitude differences and emphasize waveform morphology. A 10-fold cross-validation scheme was used to validate the results. For each fold and each channel, the training set was used to generate a representative waveform template for both classes by computing a trimmed mean of 20% (removing the 20% most extreme values) of the trials belonging to each class. This approach reduces the influence of outliers or noisy trials, providing a more robust estimate of the mean ERP. During the testing phase, each trial was compared with the two templates using the correlation distance $(1 - r_{xy})$, where r_{xy} is the correlation between the signals x and y computed as:

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (4.5)$$

and where \bar{x} and \bar{y} represents the mean of the two signals and N is the number of samples (in this case $N = 154$). This is a metric relatively insensitive to offsets and amplitude differences, instead focusing on the waveform morphology. The main classification metrics were then calculated for each channel: accuracy, precision, recall, specificity, and F1-score.

4.3 Time Frequency Feature Extraction and Classification

Given that ERPs do not necessarily provide enough information to allow trials to be classified into different phonetic families, the attempt to classify *OV* vs *V** syllables was repeated in time-frequency. The aim in this case is to identify the frequency bands and channels that best help to discriminate between the two classes and to use this information to gain insights into SI activity. The preprocessing used is the same as shown in Figure 4.3 with the exception of filtering: the passband in this case was of [4-100] Hz. The classification procedure followed the approach used by Moon (2022) [30], where the time-frequency features are extracted using the Continuous Wavelet transform (CWT). First, the data were zero-padded at both the beginning and the end to avoid edge effects. For each channel and trial, the CWT was computed using a complex Morlet wavelet, with 12 voices per octave and a frequency range of [4-70] Hz. The resulting wavelet coefficients were then saved and divided into two classes according to the distinction described previously (*OV* vs *V**). To select significant features, a cluster-based permutation test was performed. In this procedure, for each time-frequency pixel a *t*-test was applied to assess whether there was a statistically significant difference ($p > 0.05$) between the two considered classes. Cluster of features were defined as groups of at least 4 adjacent suprathreshold pixels (4-connected clusters). The cluster statistic was computed as the sum of the *t*-values of the cluster components, also known in the literature as the cluster mass. This value was then evaluated using a permutation test: trial labels were randomly shuffled $N_{perm} = 1000$ times, and for each permutation the maximum cluster mass $M_{perm}^{(i)}$ was stored. The ten largest observed clusters mass M_{obs} were compared against the null distribution to obtain a corrected *p*-value as follows:

$$p_{corr} = \frac{1 + \sum_{i=1}^{N_{perm}} (M_{perm}^{(i)} \geq M_{obs})}{N_{perm} + 1} \quad (4.6)$$

Only clusters with significant corrected *p*-values were retained. This analysis ensures that the selected features are statistically reliable and removes isolates features that might reach significance by chance. For each significant time–frequency feature, both the real and imaginary parts of the wavelet coefficients were stored as independent features. Subsequently, these were selected using the Minimum Redundancy Maximum Relevance (mRMR) algorithm, retaining a maximum of 20 informative features per fold in cross-validation. This algorithm orders and selects the features by maximizing the rate between their relevance and redundancy, both com-

puted using the Mutual Information as a metric, to maximize the predictive power of the selected set of features. Then a Support Vector Machine (SVM) classifier with a radial basis function (RBF) kernel and automatic standardization was trained. Validation was performed using a 10-fold cross-validation, and for each fold, the following metrics were computed: accuracy, precision, recall, and F1 score.

Chapter 5

Results

This chapter presents the results of the analyses. First, the scalp distributions of the mean ERPs in Grand Average and at the subject level are shown, with the aim of providing an overview of the shared ERPs and inter-subject variability. The results of the unsupervised clustering analysis are then reported through contingency matrices that illustrate the relationship between the identified clusters and the phonetic categories of the syllables. Finally, the results of the template matching classification procedure are presented. No paragraph is dedicated to the results of the time-frequency classification, as this analysis did not produce statistically significant results for any of the subjects tested.

5.1 Grand Average ERPs visualization

The topoplots obtained from the grand average analysis are shown in Figures 5.1, 5.2 and 5.3.

The figures show little difference between the classes compared in the three types of contrasts (syllabic structure, consonant family, vowel family). The areas where a waveform seems to stand out most are the parietal, occipital and frontal regions. In the parietal and occipital areas, large peaks are visible at the same latency. In the frontal area (*AF3*, *AF4*, *F5*, *F6* channels), a negative peak with a latency of approximately 150 ms is visible. The ERP images displayed to analyze the representativeness of the average in the trials show a consistent trend in the frontal and occipital channels. The ERP images of channel *AF3* (Figure 5.4) and *O1* (Figure 5.5) are shown below as representative of the activity of the frontal and posterior regions, respectively.

Both figures show high alignment of trials and no visible trend due to syllables categories. The waveforms shown in Figure 5.4 show, in particular, a marked negativity at a latency of 100 to 120 ms (N100) followed by a positivity at 250 ms (P250). Similarly, in the temporal and occipital regions, (Figure 5.5) shows the following peaks: N120, P200, N350. Few differences in categories are visible in the temporal and central areas.

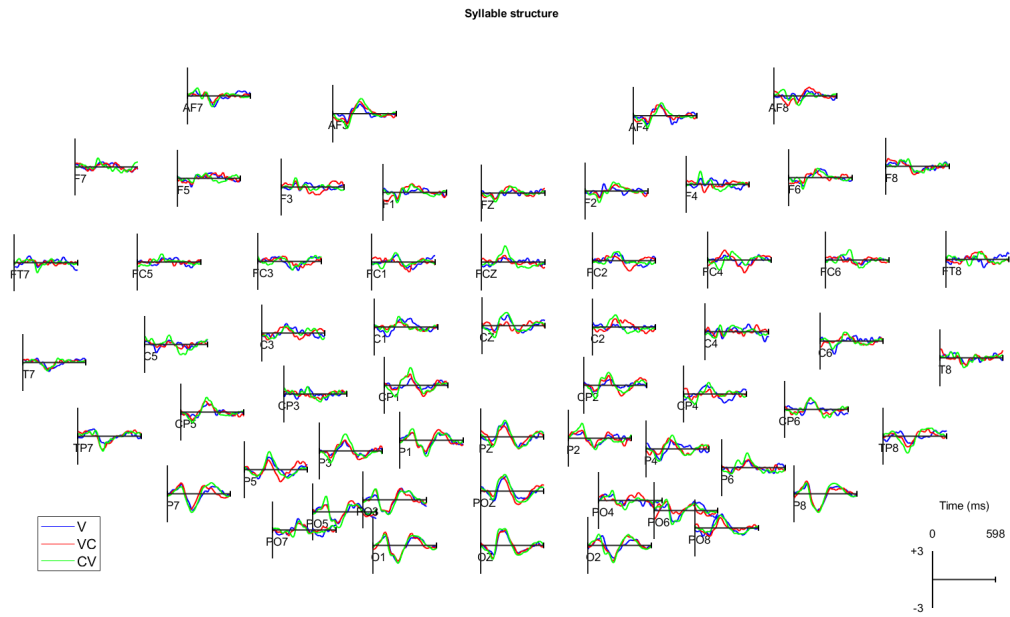


Figure 5.1: Grand average ERP topoplots across syllable structures: V - vowels only; VC - syllables starting with vowels; CV - syllables starting with consonants.

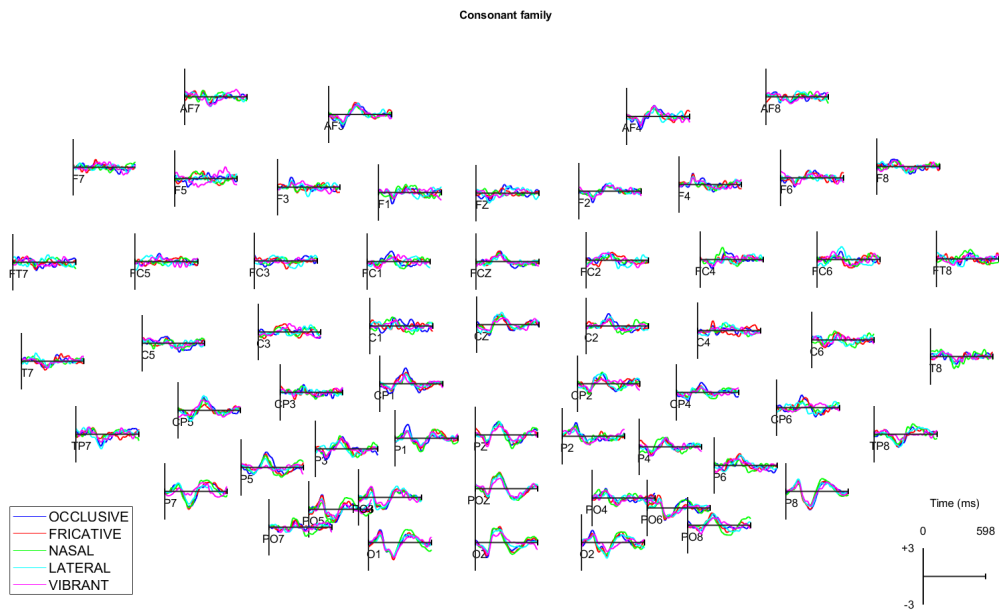


Figure 5.2: Grand average ERP topoplots for syllables beginning with consonants, grouped by consonant phonetic families.

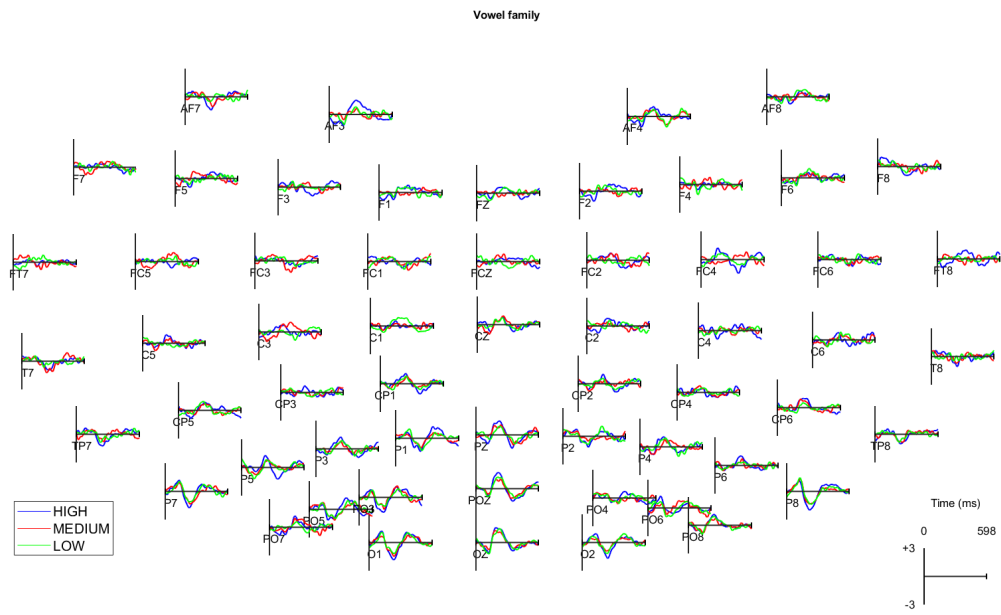


Figure 5.3: Grand average ERP topoplots for syllables beginning with vowels, grouped by vowels phonetic families.

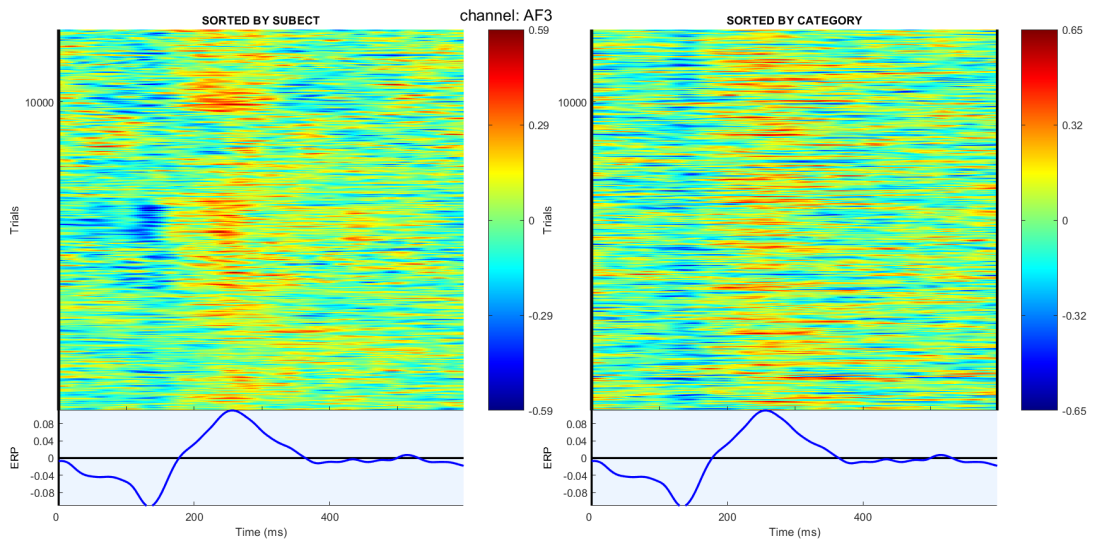


Figure 5.4: ERPimage channel *AF3*, all trials. On the left plot, trials are sorted by subject; on the right, they are sorted according to the syllable code, which follows the sonority-based classification described in Subsection 4.2.2.

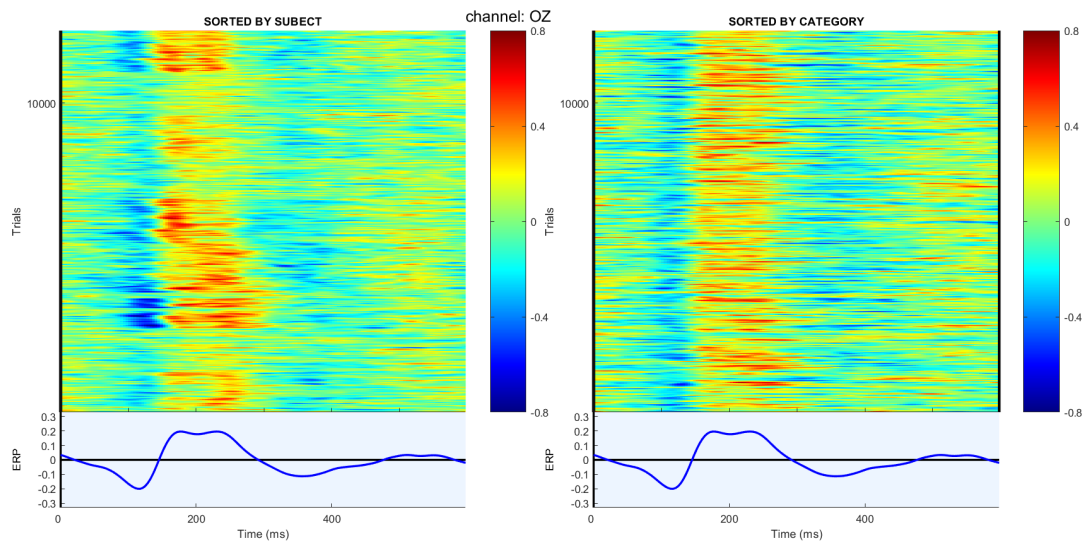


Figure 5.5: ERPimage channel *OZ*, all trials. On the left plot, trials are sorted by subject; on the right, they are sorted according to the syllable code, which follows the sonority-based classification described in Subsection 4.2.2.

5.2 Subject level visualization

At the subject level, waveforms emerge that are not seen at the Grand Average level. In particular, slight differences in waveforms between the various syllabic structures seem to emerge in the parietal and temporal regions. As an example, Figure 5.6 shows the ERP scalp map of subject *f9*.

5.3 ERP clustering results

No channel-cluster number combination yielded a significant chi-square value either in the Grand Average or at the individual subject level. For illustration purposes, Figure 5.7 shows the three contingency matrices obtained for the channel combination *FCZ* - $n_{\text{clusters}} = 9$. They show the frequency of the various phonetic categories within the clusters. As is evident, no code-cluster relationship stands out above the others: for each matrix and cluster, the frequency pattern of the categories is similar. The differences in frequencies within a column reflect only the structure of the syllables: the first letter of the syllable is often a plosive, the second is often a vowel (codes 7, 8, 9) and the third is often absent (code 0). Table 5.1 reports the results for this channel - cluster combination.

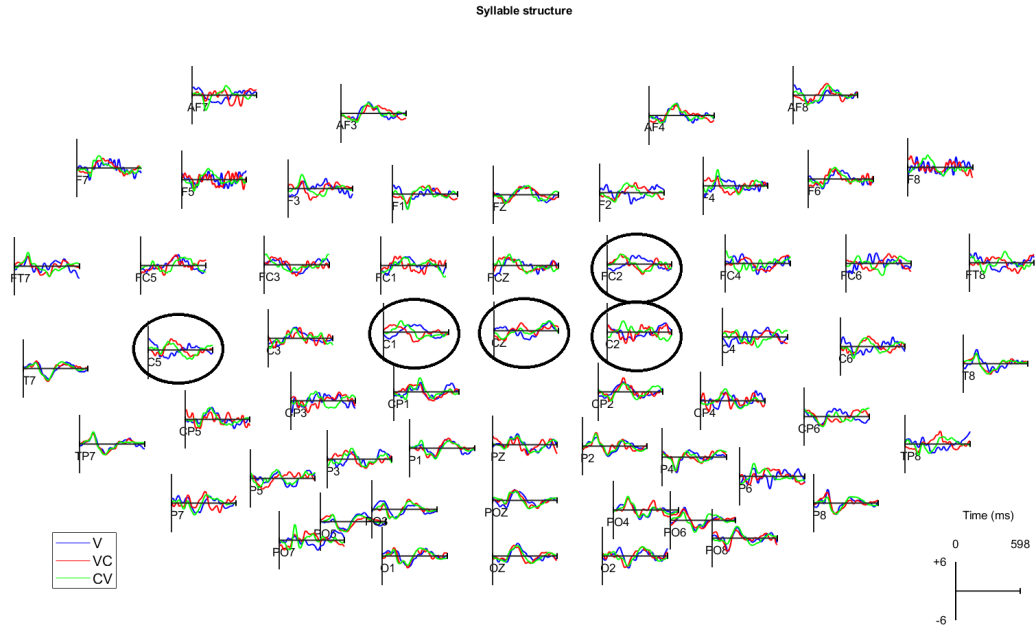


Figure 5.6: Subject-level ERP topoplot. The channels showing differences in ERP shape between the three classes represented (V, VC and CV) are circled.

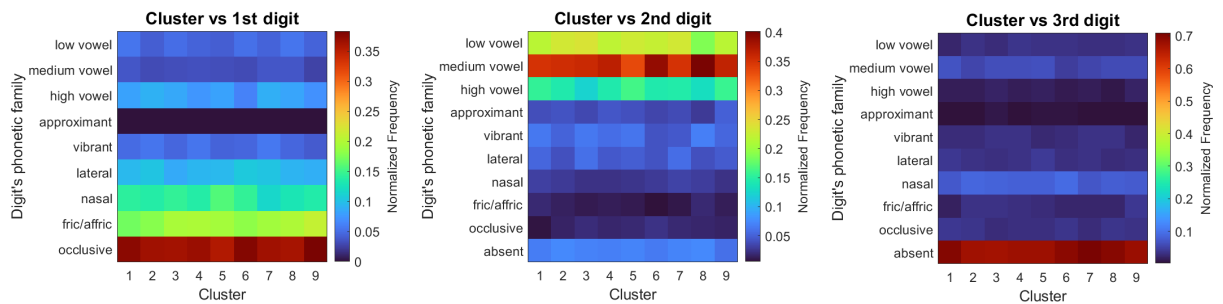


Figure 5.7: Contingency matrices in Grand Average obtained for the *FCZ* channel and using 9 clusters. The cells of the matrices show the frequencies with which syllables containing a phoneme from a certain family (y-axis) belong to the various clusters (x-axis). The frequencies are normalized by column to highlight the distribution pattern of phonetic categories within clusters.

Table 5.1: Results of the chi-squared test for channel FCZ with $n_{clusters} = 9$. The table reports the chi-squared statistic (χ^2), degrees of freedom (df), and corresponding p -values for each digit position.

Digit position	χ^2	df	p-value
2	43.93	64	0.9740
3	71.66	81	0.7615
4	54.10	81	0.9907

5.4 ERP classification results

The classification process yielded above-chance results for some channels. Accuracy was low on average, reaching peaks at the subject level of $acc = 0.5962$ (for subject $j7$ at channel $PO4$). The average accuracies reported in Figure 5.8 are even lower due to the low similarity between the channel accuracies at the subject-level. The difference in classification accuracy of the channels between subjects is shown in Figure 5.9.A: the predictive strength of the channels appears to be very scattered both at the individual subject level and at the dataset level, with no macro-areas of particular relevance standing out.

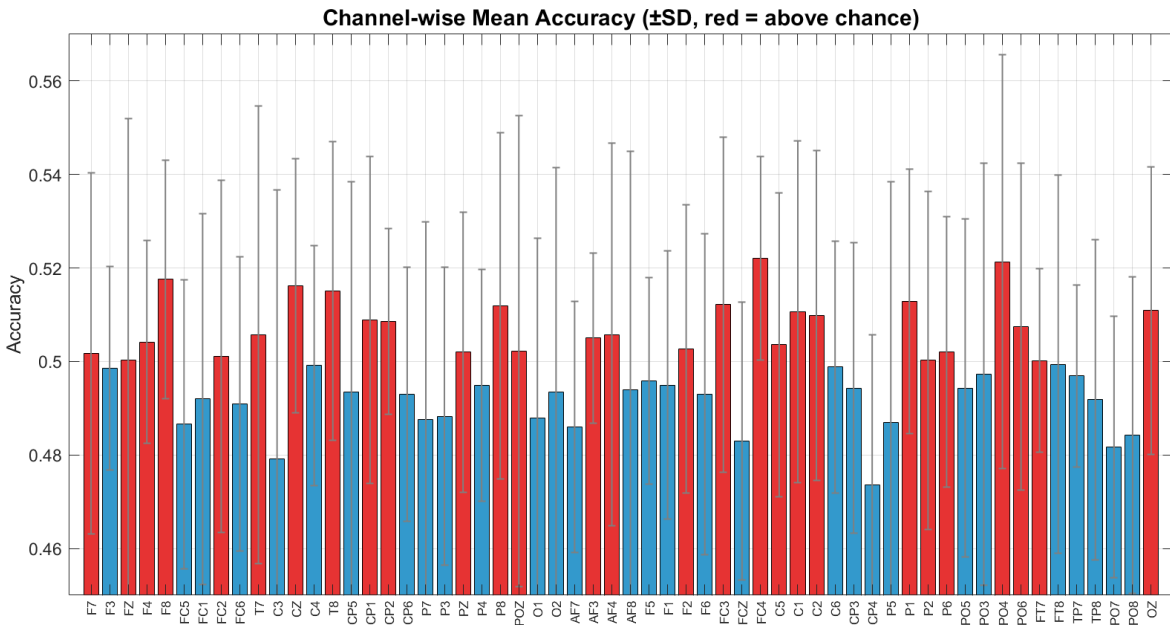


Figure 5.8: Mean classification accuracy per EEG channel across subjects with standard deviation (gray bars). Red bars indicate the channel with above-chance average accuracies, while blue bars represent below-chance average performances.

Since visualizing only the average accuracy can be misleading (SI activity could vary greatly

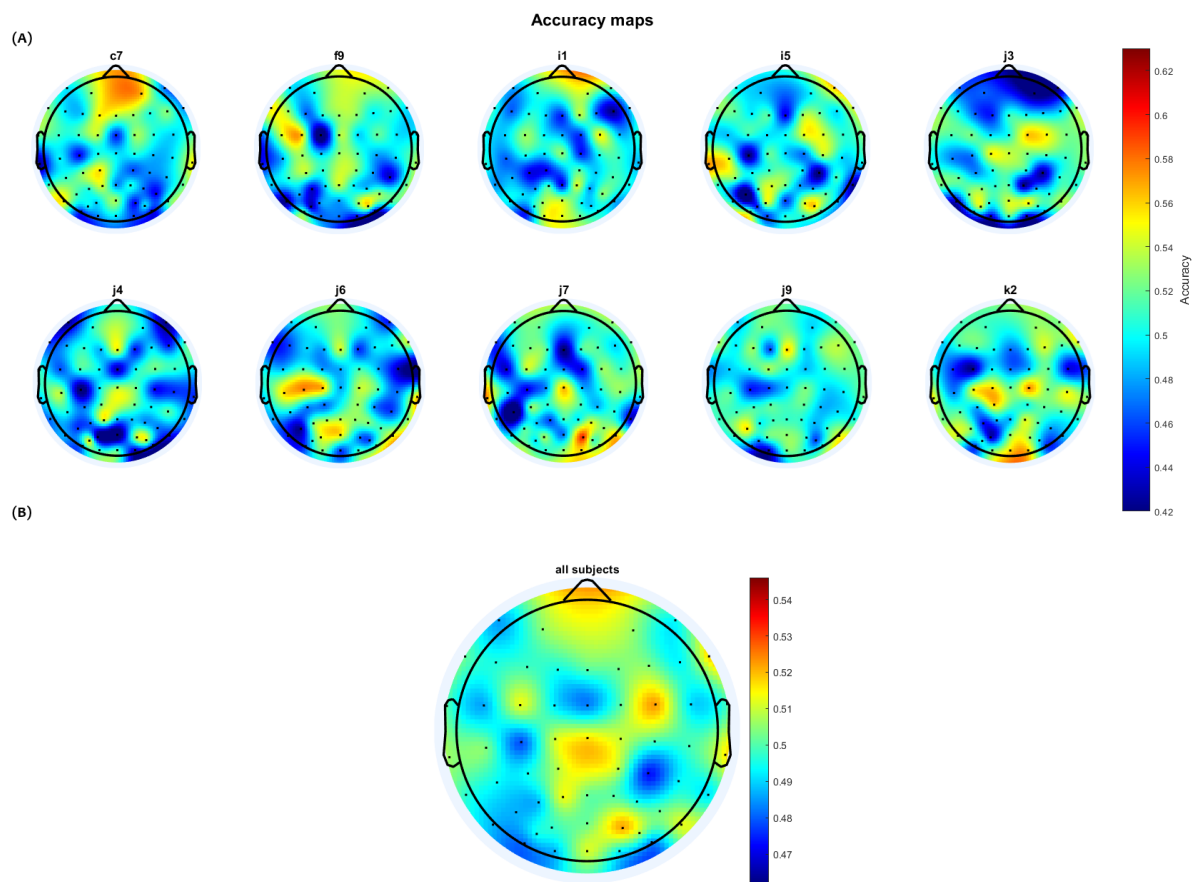


Figure 5.9: (A) Scalp heatmaps represent the accuracy obtained per channel at the subject level. (B) Mean accuracy across subjects.

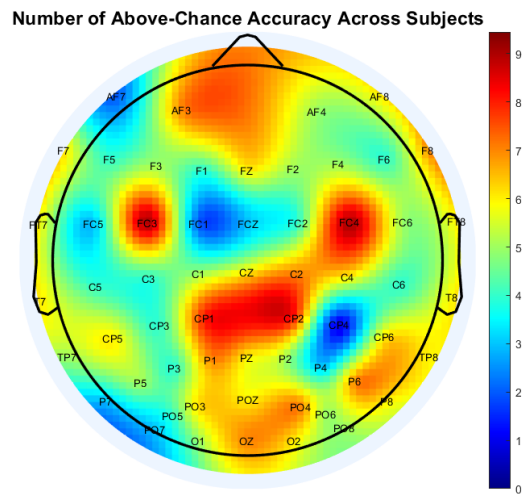


Figure 5.10: Scalp heatmap showing the count of channels with above-chance accuracy among all 10 subjects.

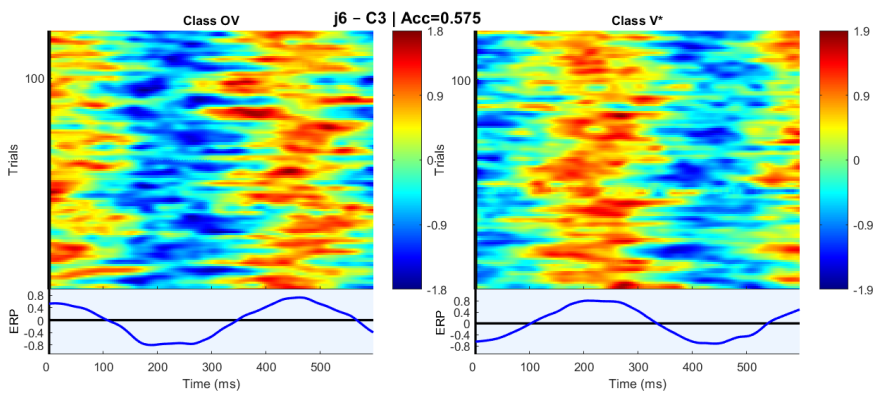


Figure 5.11: Example of correctly classified trials.

from subject to subject, and low accuracy from one subject can obscure high accuracies from another), a map (Figure 5.10) was created to display the frequency with which each channel appears significant across the entire dataset. This representation highlights how some channels achieved significant performance in more than one subject. In particular, channels *FC3* and *FC4* were significant for 9 out of 10 subjects, followed by *CP1* and *CP2* with a frequency of 8 out of 10.

Figure 5.11 shows an example of correctly classified trials (accuracy of 57%) for the subject *j6* and channel *C3*. What can be seen from this figure is how much the average of the trials is more similar to an oscillation than to an ERP with defined peaks.

Chapter 6

Discussion and Limitations

6.1 Evoked Potentials

Neither image (Figures 5.4, 5.5) representing the ERPs visible in the frontal and posterior areas, respectively, shows a relationship between the waveforms of the trials and the codes representing the phonetic categories of the syllables: no clear trend can be observed between conditions. Conversely, when sorted by subject, clusters of temporally aligned trials become visible: inter-subject differences are more pronounced than intra-subject differences related to the type of syllable produced. Both images show strong alignment and, as already mentioned, little correlation with the type of syllable produced. This suggests that the observed activity reflects evoked potentials rather than ERPs specifically related to syllabification processes in SI. Therefore, in order to interpret these waveforms, which are likely to be evoked potentials, it is necessary to refer to the literature. As regards the occipital area, these could be simple visual evoked potentials (VEPs) [52], [53] caused by screen flickering: the rapid change in syllables visual stimulus at a frequency of approximately 1.6 Hz (the duration of the trials is 600 ms). Another interpretation consistent with the area and type of stimulus is that it could be N170, a potential evoked by reading words[54]. Clearly, the problem with this interpretation is the latency of the negative peak, which in the data of this study is visible earlier (approximately 100 ms) than that reported in the literature (170 ms). Interestingly, both the literature and the data presented here suggest that, whether referring to SSVEPs or N170 components, neither appears to show any relationship with the type of syllable being read. This is actually a positive result, as it indicates that these potentials - arising from the experimental protocol rather than from speech imagery itself - will not act as confounding factors in the subsequent classification processes aimed at developing an endogenous speech BCI. However, it should also be noted that these are not potentials linked to SI activity, so they could be eliminated during the IC selection phase in future analysis. The interpretation of the frontal evoked potential is less clear. These waveforms show two fun-

damental components: N100 and P200. These components, located in the frontal area, may be related to residual P300 responses or related to N100 potentials[55]. However, these interpretations are not very valid, as the P300 wave is evoked by rare stimuli[56], which is not the case here. The N100 potential, on the other hand, is a potential associated with hearing sounds and is typically located in the frontal area. According to what is stated in Chapter 2, it could be possible that SI activity is capable of generating a quasi-perceptual experience that can even replicate the evoked potential associated with hearing a sound. However, this interpretation does not hold, as N100 is also associated with rare and loud stimuli. Furthermore, the described waveform is visible in virtually all tests and with a high degree of alignment (Figure 5.4), making it difficult to attribute it to endogenous activity. The high temporal alignment and similarity of the trials also exclude that this waveform is related to eye blink or other eye-related artifacts. The most plausible hypothesis is that these are components of evoked potentials related to the cognitive processing of visual information that comes from the occipital areas. In particular, it could play a role in integrating visual information and memory and in semantic processing: it represents a transition phase between visual processing and phonological activation, i.e. the beginning of the conversion of graphemes into phonemes. [57], [58]

6.2 Speech Imagery ERPs

As mentioned, no significant ERPs attributable to SI activity emerge from the Grand Average plot of ERPs. However, as shown in Figure 5.6, some differences in the ERP waveform can be identified at the subject level. This discrepancy between the Grand Average and the subject level could be explained in two ways: (1) syllabification activity varies greatly between subjects, so averaging flattens the waveforms; (2) alternatively, reducing the number of trials (by approximately a tenth, corresponding to the number of subjects considered) increases the influence of noisy trials, potentially leading to the appearance of spurious waveforms erroneously interpreted as ERPs. Consistently with this second interpretation the exploratory study that leverages on the unsupervised hierarchical clustering and the chi-test did not reveal any significant relationship between waveform and syllabic clustering. However, this test was exploratory in nature and trying to capture any significant relationship considering all the different syllables simultaneously may have been a limitation. The classification test, which considered only two types of syllables (*OV* syllables vs *V** syllables), reported results above the chance level. However, this results needs to be analyzed in greater depth. In addition to the significance of specific channels, it is important to consider the consistency with which electrodes *FC3*, *FC4*, *CP1* and *CP2* emerged across subjects. These findings also have plausible neurophysiological bases. Moreover, the bilateral symmetry of these electrodes further strengthens their interpretability, suggesting that the

observed effects are not the result of sparse and casual differences between classes but instead reflect a meaningful involvement of homologous cortical regions. In particular, *FC3* and *FC4* correspond approximately to the premotor regions, which are known to remain active during language imagery due to the preservation of motor planning and internal simulation mechanisms that typically precede explicit articulation. Similarly, *CP1* and *CP2* are located above the somatosensory areas, and their involvement is consistent with the generation of anticipated somatosensory consequences during imagined speech. However, despite these promising considerations, several limitations must be acknowledged. To begin with, the channels mentioned have been consistently above chance, but have not always reported the best accuracy 5.9. Furthermore, it should be noted that the accuracies obtained are very low. Although this is partly expected at the Grand Average level, given the high degree of inter-subject variability in both performance and cortical activation patterns, at the individual level the accuracies of the individual channels never exceed 60%. Such modest performance is not sufficient to argue that the activity analyzed truly reflects the SI task. Further consideration arises from the observation of ERP images (Figure 5.11), suggesting that the waveforms considered do not exhibit the typical characteristics of event-related potentials but rather resemble spontaneous oscillations. If this interpretation is correct, the few above-chance results could be attributed to simple spurious correlations between the signal phase in the syllabation event time window and the label assigned to the syllable, rather than to a real neural correlate of SI activity. However, these considerations are not sufficient to argue for the complete absence of syllabification-related ERPs. It should be noted that the experimental protocol was not designed specifically to characterize the neurophysiological correlates of syllabification during SI. The two classes used are intrinsically heterogeneous, as each contains syllables composed of distinct phonemes. Moreover, unlike studies on language processing, there is no reliable way to temporally anchor the syllabification process, or, more broadly, the SI activity itself, nor to assess its successful execution, since it produces no observable behavioral output. Another obstacle to obtaining meaningful results is undoubtedly the lack of proficiency of the subjects included in the dataset. Speech imagery is a complex task and, unlike motor imagery - where it is generally sufficient to determine the onset of the MI activity - its decoding also requires a precise temporal segmentation of the internal structure of the task. In any case, as in motor imagery, training has also been shown to improve BCI performance in speech imagery [47], so future studies that aim to characterize this paradigm could consider focusing on trained subjects. Regarding the attempt to classify the two classes using time-frequency features, the absence of significant results is likely to be attributable to the limitations described above and the simplicity of the used method rather than to a genuine lack of discriminative features. Indeed, the literature on speech BCIs based on EEG (Chapter 3) suggests that effective classification is possible, often involving more than two classes.

It is also worth noting that more sophisticated and refined approaches, particularly within the deep-learning domain, have a substantially higher potential to address challenges of this kind.

Chapter 7

Conclusions

The objective of this thesis was to characterize the activity of speech imagery. In particular, the study focused on analyzing whether potential neural correlates of SI activity carry discriminative power between two families of syllables, based on the assumption that if the neural activity changes depending on the syllable being produced, then it is likely that such activity is associated with the task under investigation. Specifically, the objective was to determine whether the syllabification process elicits well-defined ERPs as happens in speech perception, an activity that shares many underlying mechanisms with speech imagery. Any positive results would have been highly relevant for segmenting continuous speech imagery. The findings presented do not appear to support the presence of syllabification-related ERPs. The oscillatory nature observed in the trials, combined with the low classification performance, suggest that the few above chance results may reflect spurious associations between the phase of the oscillation and the trial's class rather than genuine ERP components. However, as previously discussed, this study does not rely on a dataset specifically designed for this purpose, and therefore no definitive conclusion can be drawn regarding the absence of syllabification-related ERPs in SI. Future studies may revisit these analyses using more homogeneous syllable classes, including subjects with higher proficiency in SI, and choosing experimental protocols that ensure better temporal alignment of the activity. It is worth noting, however, that both the alignment issue and the proficiency issue remain open challenges in the study of SI. In the absence of observable behavioural outputs, temporal alignment must rely on external stimuli, which inevitably introduce task-irrelevant neural activity. Similarly, user proficiency in SI is not yet a parameter that can be assessed through objective and measurable criteria. Analyses could also be extended by examining frequency sub-bands rather than the broad [1 -20]Hz range, and by employing more sophisticated classification methods. An additional improvement could involve modelling temporal dynamics at the ROI level rather than at the channel level, to capture more complex and meaningful dynamics. regarding the dataset itself, a substantial presence of protocol-related

evoked potentials was identified. In future studies using this dataset, these components could be removed during the ICA stage to avoid contamination of the SI-related activity. In conclusion, speech imagery is undoubtedly one of the most complex paradigms used for BCIs. It is complex both in its execution and in its characterization, as it does not produce any kind of observable effect. However, its characterization is a fundamental aspect in the development of tools such as EEG speech BCIs, a useful and safe resource for people who have lost the ability to communicate with the outside world.

Bibliography

- [1] S. Durga and V. Mehrotra, “Communication and its vital role in human life,” *International journal of health sciences*, vol. 6, no. S5, pp. 5940–5948, Jun. 2022, Section: Peer Review Articles. doi: 10.53730/ijhs.v6nS5.10005. Accessed: Sep. 1, 2025. [Online]. Available: <https://sciencescholar.us/journal/index.php/ijhs/article/view/10005>.
- [2] B. Geurts, “Communication as commitment sharing: Speech acts, implicatures, common ground,” en, *Theoretical Linguistics*, vol. 45, no. 1-2, pp. 1–30, Jun. 2019, Publisher: De Gruyter Mouton, issn: 1613-4060. doi: 10.1515/t1-2019-0001. Accessed: Sep. 1, 2025. [Online]. Available: <https://www.degruyterbrill.com/document/doi/10.1515/t1-2019-0001/html>.
- [3] G. Konopka and T. Roberts, “Insights into the Neural and Genetic Basis of Vocal Communication,” en, *Cell*, vol. 164, no. 6, pp. 1269–1276, Mar. 2016, issn: 00928674. doi: 10.1016/j.cell.2016.02.039. Accessed: Sep. 1, 2025. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0092867416301891>.
- [4] L. May, J. Gervain, M. Carreiras, and J. F. Werker, “The specificity of the neural response to speech at birth,” en, *Developmental Science*, vol. 21, no. 3, e12564, 2018, _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/desc.12564>, issn: 1467-7687. doi: 10.1111/desc.12564. Accessed: Sep. 1, 2025. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/desc.12564>.
- [5] K. Voity, T. Lopez, J. P. Chan, and B. D. Greenwald, “Update on How to Approach a Patient with Locked-In Syndrome and Their Communication Ability,” *Brain Sciences*, vol. 14, no. 1, p. 92, Jan. 2024, issn: 2076-3425. doi: 10.3390/brainsci14010092. Accessed: Sep. 3, 2025. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10813368/>.
- [6] M. J. Vansteensel et al., “Towards clinical application of implantable brain–computer interfaces for people with late-stage ALS: Medical and ethical considerations,” en, *Journal of Neurology*, vol. 270, no. 3, pp. 1323–1336, Mar. 2023, issn: 1432-1459. doi: 10.

1007/s00415-022-11464-6. Accessed: Sep. 2, 2025. [Online]. Available: <https://doi.org/10.1007/s00415-022-11464-6>.

- [7] S. Luo, Q. Rabbani, and N. E. Crone, “Brain-Computer Interface: Applications to Speech Decoding and Synthesis to Augment Communication,” English, *Neurotherapeutics*, vol. 19, no. 1, pp. 263–273, Jan. 2022, Publisher: Elsevier, issn: 1878-7479. doi: 10.1007/s13311-022-01190-2. Accessed: Sep. 4, 2025. [Online]. Available: [https://www.neurotherapeuticsjournal.org/article/S1878-7479\(23\)00166-6/fulltext](https://www.neurotherapeuticsjournal.org/article/S1878-7479(23)00166-6/fulltext).
- [8] M. J. Vansteensel et al., “Methodological Recommendations for Studies on the Daily Life Implementation of Implantable Communication-Brain-Computer Interfaces for Individuals With Locked-in Syndrome,” EN, *Neurorehabilitation and Neural Repair*, vol. 36, no. 10-11, pp. 666–677, Nov. 2022, Publisher: SAGE Publications Inc STM, issn: 1545-9683. doi: 10.1177/15459683221125788. Accessed: Sep. 5, 2025. [Online]. Available: <https://doi.org/10.1177/15459683221125788>.
- [9] N. Birbaumer et al., “A spelling device for the paralysed,” en, *Nature*, vol. 398, no. 6725, pp. 297–298, Mar. 1999, Publisher: Nature Publishing Group, issn: 1476-4687. doi: 10.1038/18581. Accessed: Sep. 5, 2025. [Online]. Available: <https://www.nature.com/articles/18581>.
- [10] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, “Brain-computer interfaces in the completely locked-in state and chronic stroke,” en, in *Progress in Brain Research*, vol. 228, Elsevier, 2016, pp. 131–161, isbn: 978-0-12-804216-8. doi: 10.1016/bs.pbr.2016.04.019. Accessed: Sep. 3, 2025. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0079612316300450>.
- [11] O. Bai, P. Lin, D. Huang, D.-Y. Fei, and M. K. Floeter, “Towards a user-friendly brain-computer interface: Initial tests in ALS and PLS patients,” *Clinical Neurophysiology*, vol. 121, no. 8, pp. 1293–1303, Aug. 2010, issn: 1388-2457. doi: 10.1016/j.clinph.2010.02.157. Accessed: Sep. 5, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1388245710002610>.
- [12] L. A. Farwell and E. Donchin, “Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials,” *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, Dec. 1988, issn: 0013-4694. doi: 10.1016/0013-4694(88)90149-6. Accessed: Sep. 5, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0013469488901496>.

- [13] C.-H. Han, Y.-W. Kim, D. Y. Kim, S. H. Kim, Z. Nenadic, and C.-H. Im, “Electroencephalography-based endogenous brain–computer interface for online communication with a completely locked-in patient,” en, *Journal of NeuroEngineering and Rehabilitation*, vol. 16, no. 1, p. 18, Dec. 2019, issn: 1743-0003. doi: 10.1186/s12984-019-0493-0. Accessed: Sep. 6, 2025. [Online]. Available: <https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-019-0493-0>.
- [14] N. Shi et al., “Steady-state visual evoked potential (SSVEP)-based brain–computer interface (BCI) of Chinese speller for a patient with amyotrophic lateral sclerosis: A case report,” en, *Journal of Neurorestoratology*, vol. 8, no. 1, pp. 40–52, Mar. 2020, issn: 23242426. doi: 10.26599/JNR.2020.9040003. Accessed: Sep. 6, 2025. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2324242622000110>.
- [15] F. R. Willett et al., “A high-performance speech neuroprosthesis,” en, *Nature*, vol. 620, no. 7976, pp. 1031–1036, Aug. 2023, Publisher: Nature Publishing Group, issn: 1476-4687. doi: 10.1038/s41586-023-06377-x. Accessed: Oct. 29, 2025. [Online]. Available: <https://www.nature.com/articles/s41586-023-06377-x>.
- [16] M. A. Khan, R. Das, H. K. Iversen, and S. Puthusserypady, “Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application,” *Computers in Biology and Medicine*, vol. 123, p. 103 843, Aug. 2020, issn: 0010-4825. doi: 10.1016/j.combiomed.2020.103843. Accessed: Sep. 5, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482520302031>.
- [17] T. Proix et al., “Imagined speech can be decoded from low- and cross-frequency intracranial EEG features,” en, *Nature Communications*, vol. 13, no. 1, p. 48, Jan. 2022, Publisher: Nature Publishing Group, issn: 2041-1723. doi: 10.1038/s41467-021-27725-3. Accessed: May 14, 2025. [Online]. Available: <https://www.nature.com/articles/s41467-021-27725-3>.
- [18] M.-S. Vry et al., “Ventral and dorsal fiber systems for imagined and executed movement,” en, *Experimental Brain Research*, vol. 219, no. 2, pp. 203–216, Jun. 2012, issn: 1432-1106. doi: 10.1007/s00221-012-3079-7. Accessed: Oct. 1, 2025. [Online]. Available: <https://doi.org/10.1007/s00221-012-3079-7>.
- [19] A. Jahangiri and F. Sepulveda, “The Relative Contribution of High-Gamma Linguistic Processing Stages of Word Production, and Motor Imagery of Articulation in Class Separability of Covert Speech Tasks in EEG Data,” en, *Journal of Medical Systems*, vol. 43, no. 2, p. 20, Dec. 2018, issn: 1573-689X. doi: 10.1007/s10916-018-1137-9. Accessed: Oct. 17, 2025. [Online]. Available: <https://doi.org/10.1007/s10916-018-1137-9>.

- [20] B. Parrell, A. C. Lammert, G. Ciccarelli, and T. F. Quatieri, “Current models of speech motor control: A control-theoretic overview of architectures and properties,” en, *The Journal of the Acoustical Society of America*, vol. 145, no. 3, pp. 1456–1481, Mar. 2019, issn: 0001-4966, 1520-8524. doi: 10.1121/1.5092807. Accessed: Oct. 1, 2025. [Online]. Available: <https://pubs.aip.org/jasa/article/145/3/1456/827450/Current-models-of-speech-motor-control-A-control>.
- [21] F. H. Guenther, “A neural network model of speech acquisition and motor equivalent speech production,” en, *Biological Cybernetics*, vol. 72, no. 1, pp. 43–53, Nov. 1994, issn: 1432-0770. doi: 10.1007/BF00206237. Accessed: Oct. 1, 2025. [Online]. Available: <https://doi.org/10.1007/BF00206237>.
- [22] F. H. Guenther, S. S. Ghosh, and J. A. Tourville, “Neural modeling and imaging of the cortical interactions underlying syllable production,” eng, *Brain and Language*, vol. 96, no. 3, pp. 280–301, Mar. 2006, issn: 0093-934X. doi: 10.1016/j.bandl.2005.06.001.
- [23] J. A. Tourville and F. H. Guenther, “The DIVA model: A neural theory of speech acquisition and production,” en, *Language and Cognitive Processes*, vol. 26, no. 7, pp. 952–981, Aug. 2011, issn: 0169-0965, 1464-0732. doi: 10.1080/01690960903498424. Accessed: Oct. 1, 2025. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/01690960903498424>.
- [24] J. F. Houde and S. S. Nagarajan, “Speech Production as State Feedback Control,” English, *Frontiers in Human Neuroscience*, vol. 5, Oct. 2011, Publisher: Frontiers, issn: 1662-5161. doi: 10.3389/fnhum.2011.00082. Accessed: Oct. 1, 2025. [Online]. Available: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2011.00082/full>.
- [25] X. Tian and D. Poeppel, “Mental imagery of speech and movement implicates the dynamics of internal forward models,” English, *Frontiers in Psychology*, vol. 1, Oct. 2010, Publisher: Frontiers, issn: 1664-1078. doi: 10.3389/fpsyg.2010.00166. Accessed: Jul. 29, 2025. [Online]. Available: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2010.00166/full>.
- [26] X. Tian and D. Poeppel, “Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation,” English, *Frontiers in Human Neuroscience*, vol. 6, Nov. 2012, Publisher: Frontiers, issn: 1662-5161. doi: 10.3389/fnhum.2012.00314. Accessed: Mar. 14, 2025. [Online]. Available: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2012.00314/full>.

- [27] L. Lu, M. Han, G. Zou, L. Zheng, and J.-H. Gao, “Common and distinct neural representations of imagined and perceived speech,” *Cerebral Cortex*, vol. 33, no. 10, pp. 6486–6493, May 2023, issn: 1047-3211. doi: 10.1093/cercor/bhac519. Accessed: Oct. 3, 2025. [Online]. Available: <https://doi.org/10.1093/cercor/bhac519>.
- [28] X. Tian and D. Poeppel, “The Effect of Imagination on Stimulation: The Functional Specificity of Efference Copies in Speech Processing,” *Journal of Cognitive Neuroscience*, vol. 25, no. 7, pp. 1020–1036, Jul. 2013, issn: 0898-929X. doi: 10.1162/jocn_a_00381. Accessed: Sep. 11, 2025. [Online]. Available: https://doi.org/10.1162/jocn_a_00381.
- [29] G. Hickok and D. Poeppel, “The cortical organization of speech processing,” en, *Nature Reviews Neuroscience*, vol. 8, no. 5, pp. 393–402, May 2007, Publisher: Nature Publishing Group, issn: 1471-0048. doi: 10.1038/nrn2113. Accessed: Mar. 14, 2025. [Online]. Available: <https://www.nature.com/articles/nrn2113>.
- [30] J. Moon, S. Orlandi, and T. Chau, “A comparison and classification of oscillatory characteristics in speech perception and covert speech,” *Brain Research*, vol. 1781, p. 147 778, Apr. 2022, issn: 0006-8993. doi: 10.1016/j.brainres.2022.147778. Accessed: Oct. 17, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0006899322000026>.
- [31] N. Mesgarani, C. Cheung, K. Johnson, and E. F. Chang, “Phonetic Feature Encoding in Human Superior Temporal Gyrus,” *Science*, vol. 343, no. 6174, pp. 1006–1010, Feb. 2014, Publisher: American Association for the Advancement of Science. doi: 10.1126/science.1245994. Accessed: Oct. 24, 2025. [Online]. Available: <https://www.science.org/doi/10.1126/science.1245994>.
- [32] B. N. Pasley et al., “Reconstructing Speech from Human Auditory Cortex,” en, *PLOS Biology*, vol. 10, no. 1, e1001251, 2012, Publisher: Public Library of Science, issn: 1545-7885. doi: 10.1371/journal.pbio.1001251. Accessed: Oct. 24, 2025. [Online]. Available: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001251>.
- [33] D. Poeppel and M. F. Assaneo, “Speech rhythms and their neural foundations,” en, *Nature Reviews Neuroscience*, vol. 21, no. 6, pp. 322–334, Jun. 2020, Publisher: Nature Publishing Group, issn: 1471-0048. doi: 10.1038/s41583-020-0304-4. Accessed: Mar. 14, 2025. [Online]. Available: <https://www.nature.com/articles/s41583-020-0304-4>.

- [34] A.-L. Giraud and D. Poeppel, “Cortical oscillations and speech processing: Emerging computational principles and operations,” en, *Nature Neuroscience*, vol. 15, no. 4, pp. 511–517, Apr. 2012, Publisher: Nature Publishing Group, issn: 1546-1726. doi: 10.1038/nn.3063. Accessed: Oct. 24, 2025. [Online]. Available: <https://www.nature.com/articles/nn.3063>.
- [35] E. M. Zion Golumbic, D. Poeppel, and C. E. Schroeder, “Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective,” *Brain and Language*, Neurobiology of Language 2010, vol. 122, no. 3, pp. 151–161, Sep. 2012, issn: 0093-934X. doi: 10.1016/j.bandl.2011.12.010. Accessed: Oct. 22, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0093934X11001982>.
- [36] F. M. Digeser, T. Wohlberedt, and U. Hoppe, “Contribution of Spectrotemporal Features on Auditory Event-Related Potentials Elicited by Consonant-Vowel Syllables,” en-US, *Ear and Hearing*, vol. 30, no. 6, p. 704, Dec. 2009, issn: 1538-4667. doi: 10.1097/AUD.0b013e3181b1d42d. Accessed: Nov. 5, 2025. [Online]. Available: <https://journals.lww.com/ear-hearing/pages/articleviewer.aspx?year=2009&issue=12000&article=00007&type=Fulltext>.
- [37] B. Khalighinejad, G. C. d. Silva, and N. Mesgarani, “Dynamic Encoding of Acoustic Features in Neural Responses to Continuous Speech,” en, *Journal of Neuroscience*, vol. 37, no. 8, pp. 2176–2185, Feb. 2017, Publisher: Society for Neuroscience Section: Research Articles, issn: 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.2383-16.2017. Accessed: Sep. 22, 2025. [Online]. Available: <https://www.jneurosci.org/content/37/8/2176>.
- [38] Giorgio Graffi and Sergio Scalise, *Le lingue e il linguaggio. Introduzione alla linguistica*, Italiano, Terza Edizione. Il Mulino, 2013, isbn: 978-88-15-24179-5. Accessed: Nov. 6, 2025.
- [39] N. S. Card et al., “An Accurate and Rapidly Calibrating Speech Neuroprosthesis,” *New England Journal of Medicine*, vol. 391, no. 7, pp. 609–618, Aug. 2024, Publisher: Massachusetts Medical Society _eprint: <https://www.nejm.org/doi/pdf/10.1056/NEJMoa2314132>, issn: 0028-4793. doi: 10.1056/NEJMoa2314132. Accessed: Oct. 29, 2025. [Online]. Available: <https://www.nejm.org/doi/full/10.1056/NEJMoa2314132>.
- [40] S. Khan et al., “Invasive Brain–Computer Interface for Communication: A Scoping Review,” en, *Brain Sciences*, vol. 15, no. 4, p. 336, Apr. 2025, Publisher: Multidisciplinary Digital Publishing Institute, issn: 2076-3425. doi: 10.3390/brainsci15040336. Ac-

- cessed: Oct. 29, 2025. [Online]. Available: <https://www.mdpi.com/2076-3425/15/4/336>.
- [41] V. S. Polikov, P. A. Tresco, and W. M. Reichert, "Response of brain tissue to chronically implanted neural electrodes," *Journal of Neuroscience Methods*, vol. 148, no. 1, pp. 1–18, Oct. 2005, issn: 0165-0270. doi: 10.1016/j.jneumeth.2005.08.015. Accessed: Nov. 6, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165027005002931>.
- [42] K. Solarana, M. Ye, Y.-R. Gao, H. Rafi, and D. X. Hammer, "Longitudinal multimodal assessment of neurodegeneration and vascular remodeling correlated with signal degradation in chronic cortical silicon microelectrodes," *Neurophotonics*, vol. 7, no. 1, p. 015004, Jan. 2020, Publisher: SPIE, issn: 2329-423X, 2329-4248. doi: 10.1117/1.NPh.7.1.015004. Accessed: Nov. 6, 2025. [Online]. Available: <https://www.spiedigitallibrary.org/journals/neurophotonics/volume-7/issue-1/015004/Longitudinal-multimodal-assessment-of-neurodegeneration-and-vascular-remodeling-correlated-with/10.1117/1.NPh.7.1.015004.full>.
- [43] C. S. DaSalla, H. Kambara, M. Sato, and Y. Koike, "Single-trial classification of vowel speech imagery using common spatial patterns," *Neural Networks, Brain-Machine Interface*, vol. 22, no. 9, pp. 1334–1339, Nov. 2009, issn: 0893-6080. doi: 10.1016/j.neunet.2009.05.008. Accessed: Oct. 29, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0893608009000999>.
- [44] C. Cooney, R. Folli, and D. Coyle, "Optimizing Layers Improves CNN Generalization and Transfer Learning for Imagined Speech Decoding from EEG," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, ISSN: 2577-1655, Oct. 2019, pp. 1311–1316. doi: 10.1109/SMC.2019.8914246. Accessed: Oct. 29, 2025. [Online]. Available: <https://ieeexplore.ieee.org/document/8914246>.
- [45] M. Bisla and R. S. Anand, "Chapter 11 - A comprehensive review on state-of-the-art imagined speech decoding techniques using electroencephalography," in *Artificial Intelligence in Biomedical and Modern Healthcare Informatics*, M. A. Ansari, R. S. Anand, P. Tripathi, R. Mehrotra, and M. B. B. Heyat, Eds., Academic Press, Jan. 2025, pp. 101–126, isbn: 978-0-443-21870-5. doi: 10.1016/B978-0-443-21870-5.00011-X. Accessed: Mar. 14, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/B978044321870500011X>.
- [46] S. Alzahrani, H. Banjar, and R. Mirza, "Systematic Review of EEG-Based Imagined Speech Classification Methods," *Sensors (Basel, Switzerland)*, vol. 24, no. 24, p. 8168,

Dec. 2024, issn: 1424-8220. doi: 10.3390/s24248168. Accessed: Nov. 6, 2025. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11679664/>.

- [47] K. Bhadra, A.-L. Giraud, and S. Marchesotti, “Learning to operate an imagined speech Brain-Computer Interface involves the spatial and frequency tuning of neural activity,” en, *Communications Biology*, vol. 8, no. 1, pp. 1–15, Feb. 2025, Publisher: Nature Publishing Group, issn: 2399-3642. doi: 10.1038/s42003-025-07464-7. Accessed: Mar. 14, 2025. [Online]. Available: <https://www.nature.com/articles/s42003-025-07464-7>.
- [48] *The KARA ONE database*. Accessed: Nov. 4, 2025. [Online]. Available: <https://www.cs.toronto.edu/~complingweb/data/karaOne/karaOne.html>.
- [49] P. M. Bertinetto and M. Loporcaro, “The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome,” en, *Journal of the International Phonetic Association*, vol. 35, no. 2, pp. 131–151, Dec. 2005, issn: 0025-1003, 1475-3502. doi: 10.1017/S0025100305002148. Accessed: Nov. 4, 2025. [Online]. Available: https://www.cambridge.org/core/product/identifier/S0025100305002148/type/journal_article.
- [50] J. Goslin, C. Galluzzi, and C. Romani, “PhonItalia: A phonological lexicon for Italian,” en, *Behavior Research Methods*, vol. 46, no. 3, pp. 872–886, Sep. 2014, issn: 1554-3528. doi: 10.3758/s13428-013-0400-8. Accessed: Nov. 17, 2025. [Online]. Available: <https://doi.org/10.3758/s13428-013-0400-8>.
- [51] A. Delorme and S. Makeig, “EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis,” eng, *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004, issn: 0165-0270. doi: 10.1016/j.jneumeth.2003.10.009.
- [52] Donnell Joseph Creel, “Visually evoked potentials,” en, in *Handbook of Clinical Neurology*, vol. 160, Elsevier, 2019, pp. 501–522, isbn: 978-0-444-64032-1. doi: 10.1016/B978-0-444-64032-1.00034-5. Accessed: Nov. 10, 2025. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/B9780444640321000345>.
- [53] A. M. Norcia, L. G. Appelbaum, J. M. Ales, B. R. Cottreau, and B. Rossion, “The steady-state visual evoked potential in vision research: A review,” *Journal of Vision*, vol. 15, no. 6, p. 4, May 2015, issn: 1534-7362. doi: 10.1167/15.6.4. Accessed: Nov. 10, 2025. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4581566/>.

- [54] S. Bentin, Y. Mouchetant-Rostaing, M. H. Giard, J. F. Echallier, and J. Pernier, “ERP Manifestations of Processing Printed Words at Different Psycholinguistic Levels: Time Course and Scalp Distribution,” en, *Journal of Cognitive Neuroscience*, vol. 11, no. 3, pp. 235–260, May 1999, issn: 0898-929X, 1530-8898. doi: 10.1162/089892999563373. Accessed: Nov. 8, 2025. [Online]. Available: <https://direct.mit.edu/jocn/article/11/3/235/3355/ERP-Manifestations-of-Processing-Printed-Words-at>.
- [55] L. D. Sanders and H. J. Neville, “An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers,” *Cognitive Brain Research*, vol. 15, no. 3, pp. 228–240, Feb. 2003, issn: 0926-6410. doi: 10.1016/S0926-6410(02)00195-7. Accessed: Nov. 10, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0926641002001957>.
- [56] J. Polich, “Updating P300: An integrative theory of P3a and P3b,” *Clinical Neurophysiology*, vol. 118, no. 10, pp. 2128–2148, Oct. 2007, issn: 1388-2457. doi: 10.1016/j.clinph.2007.04.019. Accessed: Nov. 11, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1388245707001897>.
- [57] G. A. Moiseenko, S. A. Koskin, S. V. Pronin, V. N. Chikhman, E. A. Vershinina, and O. V. Zhukova, “Components of Evoked Potentials in Frontal Cortex Areas Associated with Image Classification and Independent of Physical Characteristics of Stimuli,” en, *Human Physiology*, vol. 50, no. 6, pp. 559–568, Dec. 2024, issn: 1608-3164. doi: 10.1134/S0362119724701032. Accessed: Nov. 11, 2025. [Online]. Available: <https://doi.org/10.1134/S0362119724701032>.
- [58] J. Goslin, J. Grainger, and P. J. Holcomb, “Syllable frequency effects in French visual word recognition: An ERP study,” en, *Brain Research*, vol. 1115, no. 1, pp. 121–134, Oct. 2006, issn: 00068993. doi: 10.1016/j.brainres.2006.07.093. Accessed: Nov. 11, 2025. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0006899306022062>.