



INSTITUTO POLITÉCNICO NACIONAL

---

---

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

Digital Signal Processing Laboratory

Feature Extraction and Classification of EEG-based Overt  
and Imagined Speech

**THESIS**

To obtain the grade of:

**MAESTRO EN CIENCIAS EN INGENIERÍA DE  
CÓMPUTO**

P R E S E N T:

**Isaac Figueroa Aguilar**

Thesis advisors:

**Dr. Sergio Suárez Guerra**  
**Dr. José Luis Oropeza Rodríguez**



Mexico City

November 2019

---

---

## THANKS

---

Here go the thanks.

---

---

## Contents

---

<b>ACRONYMS</b>	<b>vii</b>
<b>GLOSSARY</b>	<b>ix</b>
<b>RESUMEN</b>	<b>x</b>
<b>ABSTRACT</b>	<b>xi</b>
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Problem Statement . . . . .	2
1.3 Justification . . . . .	2
1.4 Hypothesis . . . . .	3
1.5 Objectives . . . . .	4
1.5.1 General Objective . . . . .	4
1.5.2 Specific Objectives . . . . .	4
<b>2 LITERATURE REVIEW</b>	<b>5</b>
2.1 Early Work for Imagined Speech . . . . .	5
2.2 Recent Works with EEG Data . . . . .	8
2.3 Overt and Imagined Databases . . . . .	13
<b>3 THEORETICAL FRAMEWORK</b>	<b>14</b>
3.1 Electroencephalography . . . . .	14
3.2 Spectral Analysis . . . . .	16
3.3 Neuron Models . . . . .	16
3.4 Classifiers . . . . .	16
3.4.1 NeuCube . . . . .	16
3.4.2 Single Spiking Neuron . . . . .	16
<b>4 METHODS AND METHODOLOGY</b>	<b>18</b>
4.1 Data Selection . . . . .	18
4.2 Preprocessing . . . . .	20
4.3 Processing, Feature Extraction and Classification Approaches . . . . .	21
4.3.1 Vector-Based . . . . .	21
4.3.2 Spatio-Temporal . . . . .	36
4.3.3 Mixed Approach . . . . .	41
<b>5 EXPERIMENTAL FRAMEWORK</b>	<b>43</b>
5.1 Validation Methodology . . . . .	43
5.2 Parameters' Selection . . . . .	46

<b>6 RESULTS AND DISCUSSION</b>	<b>52</b>
6.1 Vector-Based . . . . .	52
6.2 Spatio-Temporal . . . . .	54
6.3 Mixed Approach Analysis . . . . .	62
<b>7 CONCLUSIONS AND FUTURE WORK</b>	<b>69</b>
7.1 Conclusions . . . . .	69
7.1.1 Data Selection and Processing . . . . .	69
7.1.2 Feature Extraction . . . . .	70
7.1.3 Experimental Framework . . . . .	70
7.1.4 Classification Results Analysis . . . . .	70
7.2 Contributions . . . . .	71
7.3 Future Recommendations . . . . .	72
<b>A DATA SELECTION</b>	<b>74</b>
<b>B FEATURE SELECTION BY SUB-BANDS CORRELATION RANKING</b>	<b>75</b>

---

---

## List of Figures

---

3.1	Neurophysiological basis of EEG generation (Figure obtained from [20]).	15
4.1	General block diagram of the methodology followed.	18
4.2	Some examples of signals from rejected samples.	19
4.3	Downsampling process applied to all signals.	20
4.4	Automatic overt speech trimming based on acoustic endpoints.	21
4.5	Wavelet decomposition tree with 4 levels.	22
4.6	Examples of Wavelet energy analysis by sub-band.	23
4.7	Signal decomposition with DWT.	24
4.8	Signal decomposition with WPT.	25
4.9	Signal framing from DWT decomposition.	28
4.10	PCA over features by wavelet coefficient sub-band.	29
4.11	Spearman correlation using DWT and 3 features.	31
4.12	Architecture for Vector-based classifiers.	33
4.13	Selected channels in [1, 2], and those selected in this work for overt and imagined speech using MLP and SVM (Original Figure taken from [3] for its adaptation in this work).	35
4.14	Encoded signal examples using SF and BSA.	39
4.15	Number of errors per generation during SSN training for overt and imagined speech examples.	42
5.1	Data used on each step of the validation methodology followed.	44
6.1	Initial and final firing activity with the NeuCube using overt speech samples.	57
6.2	Initial and final firing activity with the NeuCube using imagined speech samples.	58
6.3	Initial and final weights with the NeuCube using overt speech samples.	59
6.4	Initial and final weights with the NeuCube using imagined speech samples.	59
6.5	Initial and final NeuCube connections for overt speech samples.	60
6.6	Initial and final NeuCube connections for imagined speech samples.	61
6.7	Firing activity per sample class and misclassified samples highlighted after SSN testing for overt speech example.	63
6.8	Firing activity per sample class and misclassified samples highlighted after SSN testing for imagined speech example.	63
6.9	Accuracies of the best configurations per case for overt speech samples.	64
6.10	Accuracies of the best configurations per case for imagined speech samples.	65
6.11	Confusion matrices from best overt speech scores.	67
6.12	Confusion matrices from best imagined speech scores.	68

---

---

## List of Tables

---

2.1	Comparatives across imagined speech classification related works. . . . .	12
2.2	Comparatives across overt and imagined speech databases. . . . .	13
4.1	Samples used for class and user. . . . .	19
4.2	Binary grouping and samples per binary class. . . . .	20
4.3	Features extracted (and their identifiers) based on [4] and proposed subset highlighted. . . . .	26
4.4	Average Variance of selected components across wavelet approaches and segments using 3 features. . . . .	30
4.5	Average Variance of selected components across wavelet approaches and segments using 21 features. . . . .	30
4.6	Spearman correlation rankings per case and selected sub-bands . . . . .	32
4.7	Selected channels in [1, 2], and those selected in this work for overt and imagined speech using MLP and SVM. . . . .	34
4.8	Example of channel's selection based on the best (average) F1 scores. . . . .	35
4.9	Optimal values obtained after applying the DEA for BSA encoding: Overt Speech. . . . .	38
4.10	Optimal values obtained after applying the DEA for BSA encoding: Imagined Speech. . . . .	38
4.11	Overall accuracies for NeuCube data set parameter's selection. . . . .	40
4.12	Overall accuracies for NeuCube samples selection for grid search. . . . .	40
4.13	Grid search values used for the NeuCube. . . . .	40
5.1	Relationship of steps between Figure 5.1 and Algorithm 2. . . . .	45
5.2	Selected parameters for Vector-based classifiers. . . . .	47
5.3	Selected parameters for the Neucube. . . . .	47
5.4	NeuCube parameter values that provided the best overall accuracy and organized by set used. . . . .	48
5.5	Selected parameters for the SSN. . . . .	48
5.6	Izhikevich behaviors used in SSN DEA. . . . .	49
5.7	Optimal SSN parameters obtained from overt speech experiments. . . . .	50
5.8	Optimal SSN parameters obtained from imagined speech experiments. . . . .	50
5.9	Optimal SSN parameters obtained from imagined speech experiments using DWT. . . . .	51
5.10	Optimal SSN parameters obtained from imagined speech experiments using WPT. . . . .	51
6.1	Overall accuracies overt speech (vector-based approach). . . . .	53
6.2	Overall accuracies imagined speech, 3 features (vector-based approach). . . . .	53
6.3	Overall accuracies imagined speech, 21 features (vector-based approach). . . . .	54
6.4	Overall accuracies overt and imagined speech (spatio-temporal approach). . . . .	55
6.5	NeuCube accuracies for overt speech samples using $\pm$ Nasal data set. . . . .	55
6.6	NeuCube accuracies for imagined speech samples using $\pm$ Bilabial data set (1st parameter set). . . . .	56

6.7	NeuCube accuracies for imagined speech samples using $\pm$ Bilabial data set (2nd parameter set). . . . .	56
6.8	Error percentages from the best results using SSN for overt speech. . . . .	64
6.9	Error percentages from the best results using SSN for imagined speech. . . . .	64
6.10	Best scores obtained with $SSN_L$ for overt speech samples. . . . .	65
6.11	Best scores obtained with $SSN_L$ for imagined speech samples. . . . .	65

---

---

## ACRONYMS

---

AFR	Average Firing Rate
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
BCI	Brain Computer Interface
BPF	Band Pass Filter
BSA	Ben's Spiker Algorithm
DBN	Deep Belief Network
DEA	Differential Evolution Algorithm
DNN	Deep Neural Network
DSP	Digital Signal Processor
DWT	Discrete Wavelet Transform
DTCWT	Double-Tree Complex Wavelet Transform
CAR	Common Average Reference
CSP	Common Spatial Pattern
db4	fourth order Daubechies
ECoG	Electrocorticography
EEG	Electroencephalography
EMA	Electromagnetic Artilugraphy
EMG	Electromyography
FIR	Finite Impulse Response
FIS	Fuzzy Inference System
Fs	Sampling Frequency
FMRI	Functional Magnetic Resonance Imaging
$H_a$	Alternative hypothesis
$H_o$	Null hypothesis
HMM	Hidden Markov Model
HT	Hilbert Transform
Hz	Hertz
ICA	Independent Component Analysis
(I)FFT	(Inverse) Fast Fourier Transform
HPF	High Pass Filter
KNN	K-Nearest Neighbour
LDA	Linear Discriminant Analysis
LIF	Leaky Integrate and Fire
LPF	Low Pass Filter
LOO	Leave One Out
LR	Logistic Regression
MEG	Magnetoencephalography
MLP	Multilayer Perceptron

MODWWPT	Maximal Overlap Discrete Wavelet Transform
MW	Moving Window
NAM	Non-Audible Murmur
NB	Naive Bayes
NIRS	Near-Infrared Spectroscopy
PET	Positron Emission tomography
PCA	Principal Component Analysis
PSP	Post-Synaptic Potential
(R)MSE	(Root) Mean Squared Error
RBF	Radial Basis Function
RF	Random Forest
RWE	Relative Wavelet Energy
TBR	Threshold-Based Representation
SVM	Support Vector Machine
SNN	Spiking Neural Network
SNR	Signal-to-Noise-Ratio
SSI	Silent Speech Interface
STFT	Short-Time Fourier Transform
SRP	Speech Related Potentials
WPT	Wavelet Packet Transform

---

---

## GLOSSARY

---

Aliasing	Phenomenon where a reconstructed signal appears at a lower frequency than the original due to a non-compliance of the Nyquist theorem.
Algorithm	Set of rules that state the steps and operations to carry a particular task.
Artifact	Recorded activity that is not of cerebral origin. They can be divided into physiologic and extraphysiologic artifacts.
(digital) filter	A lineal and invariant system that separate the frequency components from its input, which is determined by its parameters selected.
EEG	Measurement of brain electrical fields via electrodes (which act as small antennas) placed on the head.
Imagined speech	Silent expression of conscious thought to oneself in a coherent linguistic form.
Overt speech	Physical articulation and production of phonemes and/or words.
Phoneme	Smallest unit of speech distinguishing one word (or word element) from another.
Realization	A particular temporal function from a random process. For EEG data is a particular channel signal.
Sample	In this work is considered the set of channel signals from a particular subject, class and trial (or repetition).
Signal	Any physical magnitude that varies in time, frequency or any other independent variable or variables. In this work will be a synonym of a realization.
Wavelet	Oscillating amplitude function that is relatively localized and limited in both time and frequency.

---

## RESUMEN

---

Principalmente para auxiliar a personas con problemas de habla, algunas investigaciones han implementado técnicas de aprendizaje máquina para reconocer el habla directamente de la actividad cerebral. La electroencefalografía es una técnica no invasiva que se ha utilizado con este propósito. En este trabajo se describen algunas metodologías seguidas para probar si es factible reconocer el habla de señales electroencefalográficas clasificando categorías fonológicas binarias a través de experimentos independientes de sujeto.

Estos experimentos se realizaron para dos actividades mentales diferentes: habla pronunciada e imaginada (pronunciación externa e interna de sonidos o palabras, respectivamente). Por lo tanto, las muestras de cada actividad mental fueron utilizadas por clasificadores de dos enfoques diferentes: 1. basado en vectores y 2. Espacio-temporales. Para el primer enfoque fueron utilizados clasificadores tradicionales, mientras que para el segundo enfoque se utilizaron clasificadores basados en neuronas pulsantes. Para estos experimentos, fueron necesarios pasos para el procesamiento de los datos y la extracción de características.

Para ambas actividades mentales (habla pronunciada e imaginada), los resultados obtenidos con el clasificador basado en una sola neurona punsante superó a los demás en todos los experimentos. Además, los mejores resultados obtenidos para las muestras del habla pronunciada fueron con el enfoque basado en vectores, mientras que para las muestras del habla imaginada fueron con el enfoque Espacio-temporal. Discusiones, análisis, y conclusiones sobre estos resultados se hacen al final de este trabajo.

---

---

## ABSTRACT

---

Mainly to aid people with speaking problems, some researches have implemented machine learning techniques to recognize speech directly from brain activity. The electroencephalography is one non-invasive technique that has been used for this purpose. In this work are described some methodologies followed to test if it is feasible to recognize speech from electroencephalography signals by classifying binary phonological categories with subject-independent experiments.

These experiments were performed for two different mental activities: overt and imagined speech (external and internal pronunciation of sounds or words, respectively). Thus, the samples of each mental activity were used by classifiers from two different approaches: 1. Vector-based, and 2. Spatio-temporal. For the first approach, traditional classifiers were used, while for the second approach were used classifiers based on spiking neurons. For these experiments, specific data processing and feature extraction steps were necessary.

For both mental activities (overt and imagined speech), the scores obtained with the single spiking neuron classifier outperformed others in all the experiments. Besides, the best results obtained for overt speech samples were with the Vector-based approach, while for imagined speech samples were with the Spatio-temporal approach. Discussions, analysis, and conclusions from these results are made at the end of this work.

# INTRODUCTION

In this chapter is introduced a general background of the problem to solve in this work. Then, the problem's implications are stated and detailed by steps to follow. Later, the expected benefits, scopes, and limits are justified. Finally, the hypothesis and objectives for this work are stated at the end of this chapter.

## 1.1 Background

BCIs are devices that allow people to command computer-specific tasks through their thoughts. In the case of the automatic speech recognition field, these technologies have been of interest mainly to develop software that could aid people with speech problems in the future. For this reason, it has been tested the feasibility to classify consciously imagined words or phonemes with machine learning algorithms in current researches. These efforts have been made by working with registers from two different mental activities: overt and imagined speech.

The registers for such mental activities can be taken with invasive or non-invasive technologies, from which EEG and ECoG are the most used, respectively, in this field. While EEG signals are recorded with a cap of electrodes placed on the scalp, ECoG signals are registered with an electrode grid placed on the brain surface. Indeed, ECoG signals are less susceptible to noise than EEG due to the records taken directly from the brain cortex. However, this and other works consider EEG registers because experiments are less complicated to replicate with new samples from different subjects.

It is stated in a standard model that simultaneous postsynaptic potentials of neural populations produce EEG signals, but as mentioned in [5], this does not explain the meaning of EEG content. For this purpose, some computational algorithms have been adapted or created to extract information from EEG signals that helps to analyze and distinguish between different mental behaviors. In the case of imagined and overt speech, these algorithms process the EEG signals and compute certain features that a classifier may differentiate between thought or pronounced phonemes or words.

Despite some proposals and experiments performed on the literature review, the classification of sounds with overt and imagined speech registers is still unsolved. This problem has been explored with few methods due to the lack of knowledge of the phenomena. For this reason, the contributions of this work, described in this thesis, are the new methods used with these data to perform and validate the classification that considers their intrinsic spatial and temporal aspects. Besides, some state-of-the-art and new classifiers are used to be compared in these experiments.

## 1.2 Problem Statement

The human brain is the most complex organ. For this reason, the development of BCIs, the recording of mental activities, and the interpretation of brain data are non-trivial tasks. Additionally, in the case of EEG data, there is a poor spatial resolution, the registers vary across different subjects and currently it is still unknown the precise identification of activated areas during imagined and overt speech recordings.

Despite these problems, technology has improved in EEG acquisition devices. Besides, some capture protocols have been created to analyze data in a controlled environment. However, from a computer science perspective, few methods have been explored to analyze and classify overt and imagined speech data. For this reason, the following general steps are identified to be covered:

- An analysis of the recorded data.
- Processing phases to minimize noise and artifacts inherent in EEG signals.
- The extraction of features able to discriminate different phonological classes.
- Classifications that consider spatial and temporal aspects of the data.

Each related work in the literature review covered some of these steps partially since currently imagined speech is a not broadly researched area.

Due to that, it emerges the necessity of using methods across all these steps. Besides, several experiments are necessary to be performed to provide reliable results.

## 1.3 Justification

Classification of overt and imagined speech is an emerging research area. Due to that, currently, it exists a few related works that use EEG data. Because of the limited number of methods used for these classification tasks, it is essential to make some actions for each involved step.

Firstly, an analysis of the data is usually not made in many works because the researchers rely on the signal acquisition of the database provider. However, since it does not exist currently databases broadly used and tested for overt and imagined speech, it is convenient to examine the data.

Moreover, the sensors (in this case, the electrodes) used to capture the signals are susceptible to noise and errors made during the recordings that sometimes are not noticed when databases are created. For this reason, all the data samples in this work were observed in time to detect and reject those that present some anomalies respecting with other EEG samples.

Then, processing steps are necessary for EEG signals to reduce the noise added by the human body or by the capture device. Besides, EEG signals are non-linear and non-stationary (i.e., the amplitudes and lengths across all samples vary over time). Due to these properties, some processing techniques have been used for EEG signals analysis. The wavelet outstands from others because it provides useful spectral and temporal information if a good selection is made.

Nevertheless, for overt and imagined speech has not been tested and compared different approaches to compute wavelets, nor founded a particular wavelet proper for such mental activities. Finding a wavelet for these mental activities is out of the scope of this work. However, for this step, two different approaches to compute wavelets have been tested and analyzed.

Next, it has been proposed certain features for EEG data in the literature review. However, they are used to characterize specific mental activities or disorders that may not be compatible with overt and imagined speech. For this reason, it is relevant to test numerous sets of features to find some that provide better characterization. In this work, the number of tested sets is limited by one set of features and a subset of the same set to find also if the dimensionality can be reduced without compromising the classification scores.

For the classification step, some classifiers that have been used in the state-of-the-art can be divided into two group approaches: those that require feature vectors as input (Vector-based) and others that extract the information directly from the EEG signals (Spatio-temporal). The problem with Vector-based classifiers used in some works is that they average the features extracted from all the channels, causing loss of information. While in other related works, all the channel features are concatenated, resulting in a vector of high dimensionality.

Working with high dimensional vectors is a problem because it requires more computation in the classifier and assumes that all the channels provide useful information for the phenomena. On the other hand, those works that use Spatio-temporal classifiers are based on deep learning, which in principle requires several amounts of samples for training and which is not possible with all the current databases available. The proposals to solve these problems are:

- Vector-based approach: Classifiers that receive as input static feature vectors from a particular EEG channel signal.
- Spatio-temporal approach: Classifiers based on spiking neurons that receive data (EEG signals, feature vectors, or encoded data) from all channels at once.

Finally, it is essential to run several experiments with the same data but with different initialization of the classifier's parameters. Then, compute statistics that provide a broader understanding of the classification results.

Performing several experiments is not usually made (or at least not reported) in the literature review, which could provide reliable results and confidence to replicate similar outcomes. For classifiers sensitive to initial parameters, several experiments were performed with the same data to report the average results in this work. Besides, it is avoided unbalanced classes to provide objective scores.

## 1.4 Hypothesis

If machine learning techniques that consider spatial and temporal data are used over imagined and overt speech registers, then it could be possible to classify phonological categories through EEG signals and provide a computational analysis.

## **1.5 Objectives**

Following are presented the objectives stated for this work.

### **1.5.1 General Objective**

Perform methods that are new in the classification of EEG-based imagined and overt speech data, and which consider spatial and temporal information.

### **1.5.2 Specific Objectives**

- Carry specific processing steps for each imagined and overt speech channel signal.
- Extract signal features suitable for each classifier.
- Perform a robust experimental framework to validate the classification results.
- Make an analysis of imagined and overt speech experiments.

---

# LITERATURE REVIEW

---

In this chapter are mentioned and described the most related works with this thesis for overt and imagined speech classification. At the beginning are described the earlier works for imagined speech. Then, some recent related works that used EEG data are briefly described. These descriptions include data acquisition, classification techniques used, databases, experimental design, and comparatives of these aspects between works. At the end of the chapter are mentioned the most recent overt speech works, which used ECoG data.

## 2.1 Early Work for Imagined Speech

Speech recognition has been a continuous increasing research field since the decade of the 70s. Then, the interest in speech recognition in inaudible environments increased for military purposes and later to find alternatives for people with speech problems. Due to that, some SSI technologies emerged [6], such as articulator capture with EMA, characterization of the vocal tract using ultrasound, digital signals transform from a NAM microphone, analysis of glottal activity using electromagnetic sensors, and surface EMG of articulator's muscles or the larynx.

Later, with the advances of technology and the development of more sophisticated BCIs, the study of speech through mental activity started at the end of the 90s. Thus, the research described in [7] was the first attempt to recognize words from EEG and MEG data. This work was carried with 100 trials from 7 imagined words produced by 7 subjects. Hence, the signals were recorded with two different technologies: **1.** 16 grid EEG sensors and **2.** a 64 channel Neuroscan cap.

Moreover, three different experimental conditions were performed, in which the average of some trials, used as training samples, was computed per channel to create prototypes that later would be compared with the test samples. In both sample types, the FFT was applied, followed by a Butterworth filtering and an IFFT. Hence, the classification step consisted of comparing the distances (MSE) between the processed test samples and each class prototype per channel. Then, the sample was associated with the class, with which the least MSE was obtained.

According to a hypothesis test, the results obtained in [7] were significant but with many variabilities across the subject's data accuracies. Nevertheless, in this preliminary work was conclude that imagined speech recognition is feasible with simple experimental conditions.

Besides, an open question was made at the end of that work, which questioned the practicality of machine learning techniques dealing with Spatio-temporal data. Hence, this open question has been the motivation of further works.

Later, to develop an intuitive communication system, it was carried a research thesis in [8], from which EEG signals were recorded with a 16 channel Electro-Cap. Besides, the samples came from several trials of 5 different words and sentences databases produced by 6 subjects. However, the number of trials used for imagined speech experiments were 5 per class.

Moreover, the processing step in [8] consisted of computing the STFT over signal's overlapped segments (some experiments related to the length's and shift's selection also were done). Also the first ( $\Delta$ ) and second ( $\Delta^2$ ) derivatives of the resulting coefficients, as well as  $\Delta$ 's average ( $\bar{\Delta}$ ), were computed. Then, these features were concatenated across the channel's segments and reduced to 16 with LDA.

Next, for the classification step were built continuous HMMs (also experiments with several amounts of states and Gaussians were performed) per class, and the input samples were associated with the class from which a maximal likelihood was obtained. Besides, a LOO validation method was used to compute the classification scores and several experiment's types were carried.

Hence, subject-dependent (SD) and subject-independent (SI) experiments were performed. SD means that classifiers are trained and tested with samples of the same subject, while SI refers when classifiers are trained with samples of one or more subject's samples and tested with other subject's samples.

In the case of SI experiments, the accuracies were close to the chance. While for SD experiments, good results were achieved (with 60% from the best average accuracy). For this reason, it was concluded in [8] that the data employed were SD.

Besides, the signals were captured in different sessions with the same subjects. Due to that, other SD experiments were carried that consisted of training the classifier with samples of one session and testing it with samples from another session. The results from these experiments showed that the data in this work were also session-dependent.

As final contributions from that work, some other experiments were carried to find the channel's contributions in the classification accuracies. Thus, these experiments consisted of leaving some channels out and performing the classification with the rest channels. The hypothesis was that the lower the accuracy, the more the channels contributed to the phenomenon.

*F7* and *T5* channels are localized in the Broca's and Wernicke's areas, respectively. Broca's area is responsible for the pronunciation fluency, while Wernicke's area for the semantic processing. Following the proposed hypothesis, accuracies obtained without the *F7* and/or without *T5* channel were compared with the accuracies provided by the rest channels.

Although worse results were obtained with *F7* and *T5* channels, the accuracies using just these channels did not significantly increase compared with the accuracy obtained with all channels. However, these experiments reflected that just the channels (7 more precisely) around the motor cortex, Broca's and Wernicke's areas were enough for imagined speech recognition.

Then, in [9] were questioned the results obtained in [8]. The hypothesis to reject those results was that such performances might have been overestimated due to temporally correlated artifacts in the signals. This argument was stated since the words in [8] were presented in blocks.

Due to that, the words in [9] were organized in the following different orders: blocks, blocks reordered, short blocks, randomized, and sequential.

Moreover, the processing, feature extraction, and classification steps were similar to those in [8]. However, features were computed with the DTCWT instead of using STFT. Hence, the signals were captured with a 128 channel Electro-cap, from which 16 channels were preselected due to the amplifier's limitations. Besides, the experiments were limited by using just the SD single session approach.

Thus, the block mode (same as in [8]) provided the best average score with 45.5% of accuracy, while others achieved the chance level. For this reason, in [9] was concluded that temporally correlated artifacts indeed superimposed the signals when words were presented to subjects by blocks.

With this last work, it is noticed the difficulty and relevance involved in imagined speech data acquisition. Besides, it was suggested in [9] some improvements for this research area, which included the use of words with semantic meaning, lengths normalization, recognition feedback to the subject, and more trials per class. However, this last aspect was not covered by the following works since their main objective was to explore the signal's characteristics with simple experiments.

The work mentioned in [10] was the first in analyzing the signals by spectral sub-bands. The signals were acquired with a 128 Channel Sensor Net, from which 18 were rejected. Besides, the database was composed of 120 trials from 4 subjects, who imagined 2 syllabic classes in 3 different rhythms. Each rhythm represented the  $\alpha$ ,  $\beta$ , or  $\theta$  well-known EEG sub-bands (described in the next chapter). Hence, the 6 resulting combinations of classes and rhythms were defined as conditions, which the classifier had to recognize.

Furthermore, contrary to the other previous mentioned works, it was performed a preprocessing step in [10] by an LPF to remove mean and linear trends. Also, it was used thresholds to suppress components of high energy since these signal's adequations would aid in improving the classifier's recognition. Then, the signal's processing consisted of extracting their envelopes with the HT, which was the base for the classification.

The classifier consisted of 6 filters associated with each condition, and which were obtained with the pseudoinverse of the envelope's averages of their corresponding training signals from the same channel. Next, each test signal's envelope was extracted and compared with all filters by their inner product, with which 6 measures were obtained. Then, each measure was summed across the channels, and the maximal measure was used to determine the most likely condition.

The resulting accuracies showed classification performances above the chance level for the three sub-bands, particularly in the  $\beta$  sub-band with an average accuracy of 74.25% across the subject's scores (these experiments were SI). However, further spectral analysis over the channel's distribution was done by subject, which was a useful tool to visualize variability between trial-by-trial samples. Due to that, it was concluded that averages across the subject's data were meaningful.

In the same year, it was carried in [11] a work similar to motor imaginary works with BCIs. Due to that, SRPs were localized by grand averaging samples in time from channels associated with the motor cortex area (C3, Cz, and C4) to test if imagined speech resembles real speech movements.

The data collection consisted of samples produced by 3 subjects, who imagined 2 different vowels or did not produce any action. Hence, 50 trials from each of these 3 classes were captured with a 64 channel BioSemi ActiveTwo system and a sampling rate of 256 Hz. Although all channels were used for classification tasks, just three were employed to visualize the SRPs.

All samples were preprocessed with a BPF (1-45 Hz), and features were extracted with the CSP method. For this reason, classification experiments were performed with each pair of classes using an SVM (RBF). Besides, each experiment was SD and repeated 20 times since the training and test sets were composed of different samples on each iteration. This experimental approach provided statistics related to the classifier's capabilities, which was implemented by first time for imagined speech.

Hence, the average accuracies obtained across binary classifications and subjects ranged from 68% to 78%, which were above the chance. Furthermore, SRPs were localized with a negative trend followed by a positive shift (at approximately 300 milliseconds later) using the motor cortex channels for both vowel classes, while for no-action class, there was no apparent SRP, which was the expected behavior.

These results, supported by the symmetric patterns observed in the spatial feature's topographic images of such channels, showed that imagined speech mechanism was similar to real speech. Also, the authors envisioned their system as a natural and intuitive control method for EEG-based SSI. Due to that, recent researches focus on overt speech decoding with ECoG signals, which are briefly described in the last section of this chapter.

## 2.2 Recent Works with EEG Data

In general, the previous works described in the last section concluded the following:

- Imagined speech recognition is feasible with simple experimental conditions [7].
- Channel's selection is possible without compromising the accuracy [8].
- Classifiers might learn from artifacts rather than mental activity due to how the data was acquired [9].
- Better accuracies could be obtained with sub-bands decomposition. However, a more in-depth analysis is necessary to validate or reject that statement [10].
- Imagined speech can resemble movement-related potentials associated with real speech movements [11].

Furthermore, each of these works agreed about the high variability across the subject's and session's samples, making SI classification a challenging task. Due to that, most of the works described in this section carried SI experiments considering the methods used in the works previously mentioned.

Imagined speech classification was performed in [12] with a database that consisted of 5 words in Spanish imagined by 27 subjects 12 times each. These signals were collected with a 14-channel Emotiv Epoc cap and an Fs of 128 Hz. The hypothesis in such work was that samples from words semantically related could provide higher scores than chance.

Due to that, different methods were used to test that hypothesis. Firstly, a CAR method was employed to enhance the SNR of the EEG signals. Then, a time-scale decomposition of each signal was obtained with the DWT using a Daubechies db2 mother wavelet, with which 5 detailed and one approximation coefficient sets were obtained (a description of wavelet decompositions is made in the next chapter).

Later, it was computed the RWE from each coefficient set (from each signal) except for the D1 coefficient's sets, which were rejected. These features were concatenated across the channels, with which a feature vector was built. The dimensions of that feature vectors varied due to the two different channel's set used: **1.** a set of 5 channels close to the Broca's and Wernicke's areas, and **2.** the 14 channels.

Furthermore, SI classification experiments with both channel's set were performed. The classifiers used were: NB, SVM (linear), and RF. In the particular case of SVM, a one-versus-the-rest method was used to classify multiple classes. Hence, 10-fold cross-validation was carried for these experiments.

In the first set of experiments (using a subset of 5 channels), the best average accuracy obtained across the subject's experiments was 44.43% with the RF classifier. Due to that, just the RF classifier was used in the second set of experiments, from which the contribution rate was also reported. The contribution rate consisted of dividing the 4-channel's accuracy by the 16-channel's accuracy. Besides, the best average accuracy obtained was 60.11% by using all channels. This result outperformed the 5-channel's accuracy, with which 47.93% was achieved.

Although the best accuracy was achieved with all channels (**2.**), the contribution rates showed that the 5 channel's considered in **1.** contributed in **2.** at least with a 50% of the classification. Due to that, it was suggested in [12] that in future works would be necessary to explore the channel's contributions to reduce the amount of data without compromising recognition performance. Besides, it was concluded that indeed words related semantically aided to obtain scores above the chance.

As a continuation of the work in [12], it was explored the channel's contribution in the work of [13] with the same database and similar **a)** processing, **b)** feature extraction and **c)** classifiers. The differences or enhancements were the followings: **a)** a biorthogonal 2.2 wavelet mother used for DWT, **b)** four statistics features were computed additionally to RWE, **c)** just RF classifier was used.

Moreover, the channel's selection consisted of two stages: **1.** a Pareto's front, which was approached as a multi-objective (Wrapper) optimization problem dealing with the error rate (accuracy maximization) and the number of channels; **2.** single channel's set selection from the front, applying several FISs.

Due to these experiments, the best average accuracy obtained across the subjects was 68.18%. Hence, the configuration that provided this accuracy consisted of a FIS with 3 membership functions for both input variables (error rate and the number of channels) and approximately 7 channels. These results provided statistical evidence that the proposed method improved the scores with 50% fewer channels used in [12].

Later, a further work in [14], that used the same database as in [12] and [13], implemented a bag of features, which consisted of generating codewords based on vector quantization. Besides, a Butterworth's LPF, signal's segmentation, histograms, and a multinomial NB classifier were used.

Furthermore, a genetic algorithm was employed to seek the configuration of 4 parameters that provided the best accuracies across SD classification experiments. These parameters were the feature's type (FFT or DWT Daubechies db4 coefficients), window size ( $8 \leq W \leq 128$ ), window sliding ( $8 \leq M \leq 128$ ), and the number of clusters ( $K \leq 1000$ ). Due to that, the best parameter's set obtained consisted of FFT coefficients,  $K = 75$ ,  $W = 40$ , and  $M = 8$ .

Moreover, the classification experiments (SD) were composed of 3 approaches: **1.** considering each segment as independent, **2.** concatenating the segments by channel, **3.** considering spatial and temporal aspects with n-grams. Also, these approaches were tested with feature vectors and raw EEG signals. From all these experiments, the best average accuracy obtained was  $68.93\% \pm 12.43\%$  with raw EEG signals and using the classification approach **2**.

Finally, the methodology followed in [14] was implemented with two other imagined speech databases. Nevertheless, the accuracies were lower than the previous related work and lower than chance in the case of using that in [15] and [1], respectively. Due to that, it was concluded that methodologies could not be generalized and a more in-depth analysis must be carried by database used.

Other contemporary work in [1] described an ambitious database composed of samples coming from 8 subjects (4 were discarded), who imagined and pronounced (overt speech) 12 times (trials) 7 phonemic/syllabic and 4 words. These signals were obtained with a 64-channel Neuroscan Quick-cap and an Fs of 1000 Hz. Also, the purpose of that work was to create a database to understand multimodal relationships. Due to that, acoustic signals and facial features (from images) were also captured.

The acquisition of each sample in that work consisted of 4 phases: **1.** 5 seconds of rest, **2.** a stimulus state (visual and auditory) followed by 2 seconds of articulator's position, **3.** 5 seconds for imagined speech and **4.** a speaking state in which overt, acoustic and image signals were recorded. Hence, the database contains all these data and their corresponding sample's intervals in the case of EEG signals.

In the preprocessing step were included the removal of ocular artifacts using blind source separation, a BPF ranging from 1 to 50 Hz, mean subtraction of each channel, and a small Laplacian filtering using the neighborhood of adjacent channels.

Next, each preprocessed channel signal was divided into segments of 10% of the total length, from which each pair had an overlap of 50% of their lengths. Then, 21 features, their  $\Delta$ 's, and  $\Delta^2$ 's were extracted from the sample points of each segment. Besides, as each feature vector was composed of 1197 values, features were ranked by their Pearson correlations to reduce their dimensionalities. Thus the accuracies reported in [1] were obtained by using 5 of these features.

On the other hand, a channel's selection was based on the Pearson correlations between acoustic and imagined speech feature vectors (composed of 1197 features) from each channel. Then, the top 10 highest absolute correlations were selected, which most of them corresponded to the area's responsible for speech planning.

Moreover, SI classification experiments were performed with a LOO validation and 5 different pairs of binary sets (the 11 classes previously mentioned were grouped into these sets phonologically). These binary set pairs were: **1.** vowel-only vs. consonants (C/V), **2.** presence of nasal ( $\pm$ )Nasal, **3.** presence of bilabial ( $\pm$ )Bilabial, **4.** presence of high-front vowel ( $\pm iy/$ ) and **5.** presence of high-back vowel ( $\pm uw/$ ).

Besides, these binary sets were used in six modalities: EEG-only, facial features only, audio (acoustic) only, EEG and facial features, EEG and audio features, and all modalities. Thus, the classifiers used for these binary sets and modalities were a DBN and an SVM with two different kernels (quadratic and RBF).

The results reported in [1] belonged to average accuracies across all modalities using C/V and  $\pm/uw/$  binary sets per classifier and subject data. Despite that an ANOVA showed variability in the results across modalities with the SVM, these accuracies showed positive results for multimodal systems development with accuracies above 90% (per subject) with a DBN.

Currently, it is known that three other works have shown results using that database. It was reported in [2] the same experiments from [1] but with a more in-depth analysis of the linear feature relationships between each pair of modalities, in which skewness, variance, and the sum of the first derivate of the signal were present in the top ten feature rankings. These preliminary results envisioned future works to reconstruct acoustic features with EEG signals in a multimodal system.

Next, one of the additional databases used in the previously mentioned work of [14] was that of [1]. However, just the 4 words were used for classification using methods proposed in [14]. The accuracies varied across the subjects and, consequently, the average accuracy was below the chance. However, it was mentioned that these results were not comparable with those from [1] due to the multimodal approach. Besides, it was concluded that further analysis must be done.

Then, the same feature vectors allocated in this database were used in [16] to perform SI experiments with the 11 classes instead of using the binary categories from [1]. These features were averaged across the 10 channels selected in [1] to feed each classifier used in that work. Besides, by that time the database consisted of 14 subject samples which were employed in that work.

Furthermore, the classification experiments were two, which in general consisted of the following objectives:

1. Classification with HMMs: Experiments using normalized and non-normalized features were performed to compare their training scores obtained with continuous HMMs (same training samples were used for test). Thus, the normalization was made by using different subset configurations, with which an accuracy's improvement was obtained with normalization.
2. Feature selection: Binary classification of  $/m/$  and  $/n/$  classes were done with six classifiers: DNN, MLP, LR, SVM (quadratic), SVM (RBF), and KNN. Besides, feature selection using the Wrapper method was carried, with which feature vectors with an optimal length of 100 values were obtained. Hence, different cross-validation experiments were performed, from which, on average, the best scores were obtained with 15-fold cross-validation. Finally, a feature selection was made jointly with HMMs, and again the best scores were obtained with normalized features.

Finally, Table 2.1 summarizes the most representative works mentioned in this chapter for imagined speech classification. The enumerations **1.** and **2.** represent each experiment also enumerated in the explanations. Hence, the score's row represents the best (average) test accuracies obtained in each work across the experiments they performed, and the spatial's row means that if the work used each EEG channel data independently or not.

Table 2.1: Comparatives across imagined speech classification related works.

Subjects	7	6	21	4	3	21	8
Classes	7 words	5-11 words	5 words	2 syllables 3 rythms	2 vowels and no-action	5	7 phones/syllables 4 words
Trials	100	5	20	120	50	33	12
Language	English	English	English	English	English	Spanish	English
Channels	1. 16, 2. 64	16	128	128	64	14	62
Fs (Hz)	1. 678, 2. 500	300	300	1024	256	128	1000
Channel's selection	No	7	16	110	3 (SRP)	1. 4, 2. 14	10 (correlation)
Preprocessing	No	No	No	LPF, mean and linear trend remotion	BPF	CAR	BPF Laplacian filter
Processing	Filtering Avg. prototypes	STFT	DTCWT	HT, BPF elliptic	No	DWT Daubechies db2	segmentation
Features	No	STFT coeff. $\Delta$ , $\Delta^2$ , $\bar{\Delta}$	DTCWT coeff	HT coeff.	CSP	RWE	21 statistical $\Delta$ , $\Delta^2$
Feature's selection	No	16 (LDA)	LDA	No	No	1. 5, 2. 70	5 (correlation)
Classifier(s)	Distance MSE	HMM	HMM	Inner product	SVM	NB, RF SVM	SVM, DBN
Classification	Multiclass	Multiclass	Multiclass	Multiclass	Binary	Multiclass	Binary
Scores	34%-97%	60%	45.5%	74.25%	68%-78%	1. 47.93% 2. 60.11%	>90%
Validation	No	LOO	LOO	No	No	10-FCV	LOO
Initializations	No	No	No	No	20	No	No
Spatial	Yes	No	No	Yes	No	No	No
Experiments	SI	SD,SI	SD	SD,SI	SD	SI	SI
Reference	[7]	[8]	[9]	[10]	[11]	[12]	[1]

## 2.3 Overt and Imagined Databases

Table 2.2: Comparatives across overt and imagined speech databases.

Database	Kara One	Brainliner	SINCI
<b>Speech</b>	Overt and imagined	imagined	Overt and imagined
<b>Subjects</b>	14	3	15
<b>Classes</b>	11	3	11
<b>Class' types</b>	phones, syllables, words	vowels	vowel, words
<b>Trials (Avg)</b>	<b>1.</b> 12, <b>2.</b> 12	50	<b>1.</b> 12, <b>2.</b> 45
<b>Total Trials</b>	<b>1.</b> 1913, <b>2.</b> 1913	450	<b>1.</b> 1973, <b>2.</b> 7341
<b>Language</b>	English	English	Spanish
<b>EEG channels</b>	62	64	6
<b>Fs (Hz)</b>	1000	256	1024
<b>Device</b>	NeuroScan	BioSemi ActiveTwo	Ag-AgCl electrodes
<b>Site</b>	[3]	[17]	[18]
<b>Reference</b>	[1]	[19]	[15]

Although some works related to imagined speech were mentioned in the previous section, not all databases are available online. Due to that, Table 2.2 shows information from the three free imagined (and overt) speech databases founded. As can be seen, the information varies among the databases.

The values **1.** and **2.** in the trial's rows represent, respectively, the amount of overt and imagined samples that the databases have. For the Kara One database, these amounts are balanced between both mental activities samples. Besides, the row *Trials(Avg)* represents the number of trials (on average) generated by each subject, and the Fs showed from the Brainliner database represents the downsampled data since the original Fs was 2048 Hz. Also, the site's and reference's rows represent, respectively, the download Web page and the main work that used such database.

It can be noticed that the SINCI database has more samples than the rest. Nonetheless, fewer channels were used for the EEG recordings in the SINCI database than the others. Besides, contrary to [18], professional caps were employed for the captures in [3] and [17]. Moreover, the Brainliner database counts with the least number of samples, as well as the least amount of subjects, and it only counts with imagined speech registers.

For these reasons and due to previous works (mentioned in the last section) that employed those data, the Kara One database was selected to perform the experiments in this work.

---

# THEORETICAL FRAMEWORK

---

In this chapter are mentioned some concepts behind the methods used in this work. In the first section are boarded some generalities of EEG signals, which include a description of their generation, mental activities associated with sub-bands, and conventional data acquisition techniques. The next section encompasses introductory concepts of wavelet functions and their usage for EEG data. Later in another section, two neuron models are briefly described, which were implemented in some classifiers used in this work. At the end of this chapter, the generalities of these classifiers are also explained.

## 3.1 Electroencephalography

The electroencephalography was used in humans for the first time in 1924 by the psychiatrist Hans Berger, who suggested that periodicities in EEG signals may be associated with mental activities. Later on, at the end of the last century, many technologies were developed and enhanced to record brain activity, which includes fMRI, PET, MEG, EEG, and NIRS.

In the case of EEG, a good time resolution can be achieved with current technologies. Nevertheless, its spatial resolution is limited by the number of channels used, even though a single electrode of  $\sim 10$  millimeters covers approximately 250000 neurons [20]. Also, it is assumed that tens of thousands of synchronously activated pyramidal neurons, perpendicular to the cortical surface, generate EEG oscillations. Thus, these oscillations represent the sum of excitatory (positive) and inhibitory (negative) pyramidal neuron's PSPs transmitted among the neural network (a brief description of these process is mentioned in the neuron model's section).

Figure 3.1 depicts these EEG generation process through three different images. Image **A.** shows a coronal slice of the human brain. Then, the image **B.** depicts a zoom in the inset from **A.**, in which the scalp, skull, and cerebral spinal fluid are represented. Finally, the image **C.** shows a cortical pyramidal neuron's scheme localized in the inset from **B.** and from which the closed loops represent the extracellular current's summation.

Figure 3.1 depicts not only the EEG generation but also the challenges of acquiring these signals due to the different layers between the brain cortex and electrodes, which inevitably distorts the signals. Due to that, a uniform and homogeneous electrode's coverage on the scalp is imperative to avoid additional signal distortions. It is mentioned in [20] that electrode's distances of 2-3 cm are required to prevent distortions of the potential scalp distribution.

Due to this issue, some electrodes placement systems have been proposed, from

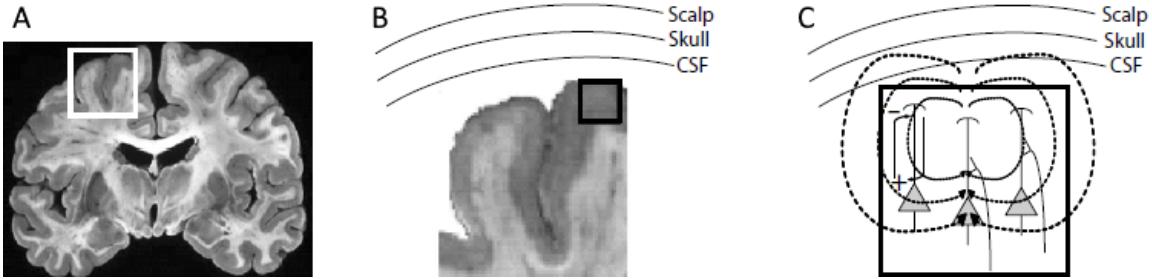


Figure 3.1: Neurophysiological basis of EEG generation (Figure obtained from [20]).

which the International 10-20 system highlights and consists of placing the electrodes at sites 10% or 20% from the total front-back (nasion-ionion) or right-left distance of the skull. Furthermore, it is also mentioned in [20] that several studies suggest at least 60 equally distributed electrodes (or channels) are required for an accurate spatial sampling of scalp activities.

Moreover, an EEG signal acquired from a particular channel represents the differential potential between that electrode and a recording reference, which ideally should be electrically inactive. Unfortunately, the reference contributes to recordings, and in some circumstances, it could be contaminated by non-encephalic activity such as muscle or electromyographic artifacts. The particular cases of artifacts caused by blinks and eye movements can be detected and corrected with additional channels that record vertical and horizontal eye movements.

According to [20], the reference choice is irrelevant for any source location; however, this is not the case for this work due to the lack of overt and imagined speech dynamics knowledge. Nonetheless, several methods have been invented for reference-dependency issues, from which average reference is commonly used and that consists of re-deriving per timestep the EEG signals against the average value across all electrodes.

Other EEG preprocessing method used to deal with these issues is the small Laplacian filter, which was mentioned in the previous chapter and was employed during the data collection in [1]. This method consists of computing the average potential difference between each electrode and the nearest four electrodes to emphasize the shallow cortical generators.

In addition to these challenges, the interactions between thalamic and cortical networks produce rhythmical activities, which are associated with particular functions, and their amplitudes typically range from 10 to 100 microvolts. These mental rhythms with their corresponding frequency band range are described as follows [20]:

- Delta band ( $\delta$ , 1-4 Hz): These oscillations are associated with sleep in healthy people and neurological pathologies, such as brain lesions and tumors. Also, they are predominant in infants during the first two years.
- Theta band ( $\theta$ , 4-8 Hz): It is prominent during sleeping. However, during wakefulness has been linked to decreased alertness, focussed attention, mental effort, and effective stimulus processing.
- Alpha band ( $\alpha$ , 8-13 Hz): This band presents amplitude differences among individuals, which are diminished by sudden alerting, mental concentration, and eye-opening. For this reason, it is associated with visual system functions. Besides, in cognitive

tasks, lower-alpha (8-10 Hz) suppression has been associated with stimulus-unspecific and task-unspecific increases in intentional demands. While upper alpha (10-12 Hz) desynchronization has been linked to sensory-semantic information processing, semantic memory performance, and stimulus-specific expectancy.

- Beta band ( $\beta$ , 13-30 Hz): It typically replaces  $\alpha$  rhythm during cognitive activity. Due to that, the  $\beta$  band has been associated with attention and vigilance states. It is also suggested that  $\beta$  increases are present during diffuse arousal and focused attention.
- Gamma band ( $\gamma$ , 36-44 Hz): These oscillations have been related to attention, arousal, object recognition, top-down modulation of sensory process, and perceptual binding (brain's ability to integrate various aspects of a stimulus into a coherent whole).

All these oscillatory rhythms have their particular aspects. However, they encompass several mental activities that are difficult to select prior experiments in a computer science fashion. For this reason, a method that provides a decomposition of EEG signals close to the bands mentioned here is described in the next section.

## 3.2 Spectral Analysis

## 3.3 Neuron Models

### 3.4 Classifiers

#### 3.4.1 NeuCube

#### 3.4.2 Single Spiking Neuron

Following the same procedure as in [21, 22], the training for the SSN consisted of the following steps:

1. An initial population is generated, which is composed of a predefined number of individuals. Hence, each individual is a vector from which each element represents an SSN's parameter. In this step, the parameter values are randomly assigned over a predefined interval.
2. The spiking neuron parameters are set with each individual's values and fed with the training samples.
3. Each training sample produced a firing activity in the neuron output, from which the firing rate of each is computed.
4. The AFR is calculated with the firing rates per associated class.
5. The third step is performed with the training and test fold data (in Chapter 5 is explained the data split), and, based on the Manhattan distance between the firing rates and the AFRs, each sample is associated with a particular class (the closest, according to the distance).
6. The individuals for the next generation are built based on the DEA and a fitness function, which consists of summing training and test fold errors.

7. The process is repeated from step 2 until the number of declared generations is completed. For all the experiments in this work, 500 generations were used.
8. The best individual was selected (the one with less number of errors) after the last generation was performed.

# METHODS AND METHODOLOGY

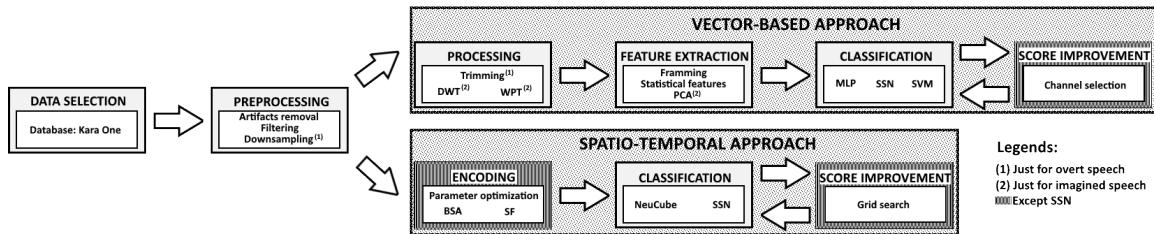


Figure 4.1: General block diagram of the methodology followed.

In this chapter is described the methodology followed in this work to classify imagined and overt speech samples. Figure 4.1 shows the block diagram with each step involved in the classification. The steps for both classifier approaches, Vector-Based and Spatio-Temporal, are differentiated with their particular methods. Besides, the methods required for the preprocessing and processing of imagined and overt speech signals are distinguished. The following sections detail the work done in each step.

## 4.1 Data Selection

The samples used in this work were obtained from the database Kara One [3], which contains EEG recordings captured with a 64-channel Neuroscan Quick cap, using the 10-20 system and sampled at 1 Kilohertz. These recordings were taken from 14 subjects and stored in individual EEGLab set files.

In addition, each set file contains, most of the cases, samples of 12 trials from 7 phonemic/syllabic classes (/iy/, /uw/, /piy/, /tiy/, /diy/, /m/, /n/) and 4 word classes (pat, pot, knew, gnaw), each. According to [1], for each class recording, the subjects were instructed to relax by 5 seconds, followed by 2 seconds of stimuli in where the participant prepare his/her articulators. Then, the participant was told to imagine speaking the phoneme or word continuously by 5 seconds, followed by a variable interval to pronounce the same class.

For this work, the corresponding 5 seconds imagined speech, as well as the following overt speech samples, were selected from the database. However, some of the preprocessed signals were discarded after observing notorious periodicities in time not seen in other revised related works. Figure 4.2 shows some examples of signals presenting this problem.

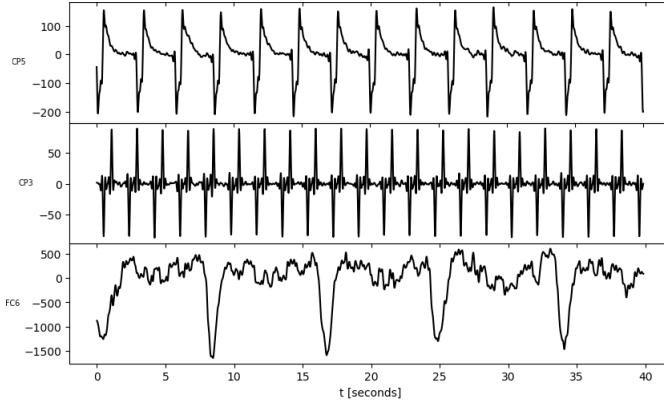


Figure 4.2: Some examples of signals from rejected samples.

Table 4.1: Samples used for class and user.

Class\Subject	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$
<i>/diy/</i>	15	8	11	9	0
<i>gnaw</i>	15	5	12	8	11
<i>/iy/</i>	15	8	12	11	6
<i>/knew/</i>	15	10	12	8	12
<i>/m/</i>	15	6	12	8	7
<i>/n/</i>	14	7	12	8	6
<i>pat</i>	15	6	12	9	12
<i>/piy/</i>	15	6	11	7	9
<i>pot</i>	15	7	12	8	11
<i>/tiy/</i>	15	7	12	8	6
<i>/uw/</i>	15	5	11	3	8
<b>TOTAL</b>	<b>164</b>	<b>75</b>	<b>129</b>	<b>87</b>	<b>88</b>

The decision over signals with that problem was out of the scope for this work. Due to that, for this work were used just some samples from the subjects *MM05*, *MM08*, *MM10*, *MM18* and *MM19* (named here as  $S_1$ ,  $S_2$ ,  $S_3$ ,  $S_4$  and  $S_5$ , respectively), that in total gave 543 samples per mental activity. Table 4.1 shows the number of samples selected by subject and class.

Also, it was followed the same binary grouping as in [2, 1]. That is, the original 11 classes were grouped into two new classes, building different binary sets based on their phonological relation. These binary sets were: vowel-only vs. consonant (C/V), presence of nasal ( $\pm$ Nasal), presence of bilabial ( $\pm$ Bilabial), presence of high-front vowel ( $\pm$ /iy/), and presence of high-back vowel ( $\pm$ /uw/).

While in [2, 1] reported results from the most unbalanced pair of sets ( $\pm$ /uw/ and C/V), in order to avoid biased results, for the experiments reported here were used the less two unbalanced binary sets:  $\pm$ Nasal or  $\pm$ Bilabial, both with 340 samples in one class and 203 in the other. In Table 4.2 are shown the samples' distribution on each phonological pair of sets, as well as the association of the original 11 classes with the new binary classes (class 1 or 2).

Table 4.2: Binary grouping and samples per binary class.

ORIGINAL CLASSES	BINARY CLASSES				
	V/C	$\pm$ Nasal	$\pm$ Bilabial	$\pm$ iy	$\pm$ uw
/diy/	2	1	1	2	1
gnaw	2	2	1	1	1
/iy/	1	1	1	2	1
knew	2	2	1	1	1
/m/	2	2	2	1	1
/n/	2	2	1	1	1
pat	2	1	2	1	1
/piy/	2	1	2	2	1
pot	2	1	2	1	1
/tiy/	2	1	1	2	1
/uw/	1	1	1	2	2
SAMPLES	94 449	340 203	340 203	352 191	501 42

## 4.2 Preprocessing

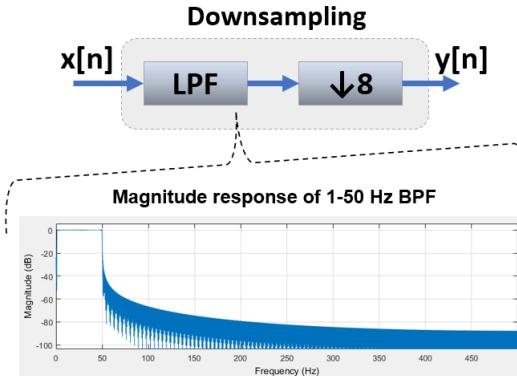


Figure 4.3: Downsampling process applied to all signals.

For this work were used the preprocessed signals from the database Kara One. Therefore, this step was the same as in [1, 2], which consisted of the following steps:

1. Removal of ocular artifacts using blind source separation with EEGLAB [23].
2. Band-pass filtering (BPF) from 1 to 50 Hz.
3. Mean values subtraction from each channel.
4. Small Laplacian filtering, using the neighborhood of adjacent channels.

Additionally, in the case of imagined speech, signals were downsampled by a factor of 8, resulting in 125 samples per second (originally were 1000). This step was necessary for these data due to the spectral analysis done with wavelets in the next step. Figure 4.3 exemplify this process. When a signal is downsampled, it is required to be beforehand filtered with an LPF to avoid aliasing.

However, as mentioned above, instead, all data were previously filtered by a BPF, which worked for this purpose. Also, the idea of downsampling was to reduce data points and reject useless frequency bands to work with the wavelets just over the bandpass 0-62.5 Hz.

## 4.3 Processing, Feature Extraction and Classification Approaches

In this section are described the processing, feature extraction, and classification steps for both approaches used in this work: Vector-Based and Spatio-Temporal. It is also described the processes for the classifier that can combine both approaches.

### 4.3.1 Vector-Based

#### Processing

For overt speech samples, this step consisted of identifying from all the time points just those that correspond when the subjects pronounced the classes. That is, to extract information just from the intervals of actual overt speech. These intervals were identified with the endpoints detection algorithm for isolated utterances proposed in [24] and broadly used in speech recognition research. Indeed, just the energy from the acoustic signals (also available in the database) was computed.

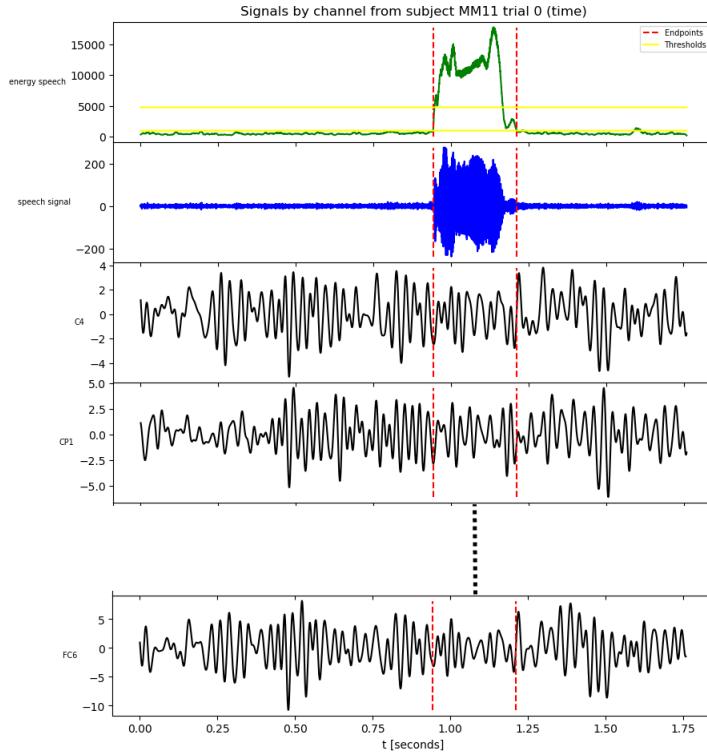


Figure 4.4: Automatic overt speech trimming based on acoustic endpoints.

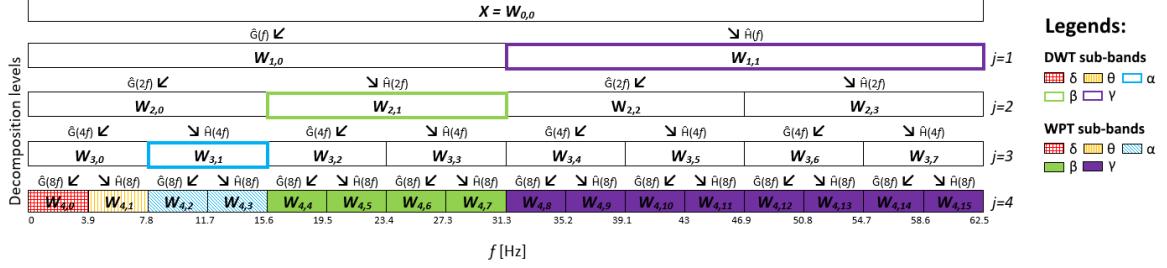


Figure 4.5: Wavelet decomposition tree with 4 levels.

Figure 4.4 shows, from the top to the bottom, the computed energy, the acoustic signal, and some EEG signals from the same sample. In the case of the energy plot, the horizontal yellow lines represent the thresholds computed in the algorithm to state the endpoints. Therefore, the vertical red dashed lines across all the plots represent the start and end of the resulting interval that was used to trim the signals.

In some cases, the algorithm failed due to noise with high energy added to the acoustic signals. That is, sometimes, the starting point or ending point was not found. For such cases, a program was implemented to select the endpoints manually by taking as reference the recordings and some plots like those from Figure 4.4.

Also, a downsampling step (neither a wavelet processing) was not performed for overt speech signals. This step was not done because the signals (after been trimmed) had few sample points that were considered just enough to compute features.

On the other hand, for imagined speech samples was necessary to process all the time points (equivalent to 5 seconds) since there were no external references to identify the exact moment when the people thought about the classes. It is just known that they were instructed to think about the classes continuously during that time interval.

The processing step consisted of computing the wavelet coefficients from the signals with two different approaches: DWT and WPT. A db4 function was used as a mother wavelet, and the coefficients were computed with the MODWPT ?? function of MATLAB, which has the advantage of making alignment in time with the original signal. Also, MODWPT avoids downsampling, which means that all resulting wavelet coefficient sets have the same length.

Figure 4.5 represents the 4 levels of decomposition applied over the imagined speech signals with DWT and WPT.  $\mathbf{X}$  represents the coefficients after correlating the EEG signal with the mother wavelet.  $\hat{G}(\lambda f)$  represents the LPF and  $\hat{H}(\lambda f)$  the HPF applied on the  $j$  decomposition level (where  $\lambda = 2^j$ ). The interval of the resulting sub-bands is shown at the bottom of the figure (in Hz). Notice that the Nyquist frequency is 62.5 Hz due to the previous downsampling step.

For both approaches, the coefficient sets  $\mathbf{W}_{4,0}$  and  $\mathbf{W}_{4,1}$  correspond closely to the  $\delta$  and  $\theta$  subbands, respectively. In the case of using DWT, the coefficients  $\mathbf{W}_{3,1}$ ,  $\mathbf{W}_{2,1}$ , and  $\mathbf{W}_{1,1}$  correspond to the  $\alpha$ ,  $\beta$  and  $\gamma$  sub-bands, respectively, resulting 5 signals to analyze. Whereas using WPT, the coefficients  $\mathbf{W}_{4,2}$  and  $\mathbf{W}_{4,3}$  correspond to the  $\alpha$  sub-band, the coefficient sets from  $\mathbf{W}_{4,4}$  to  $\mathbf{W}_{4,7}$  correspond to the  $\beta$  sub-band, and the coefficient sets from  $\mathbf{W}_{4,8}$  to  $\mathbf{W}_{4,15}$  correspond to the  $\gamma$  sub-band, resulting 16 signals to analyze.

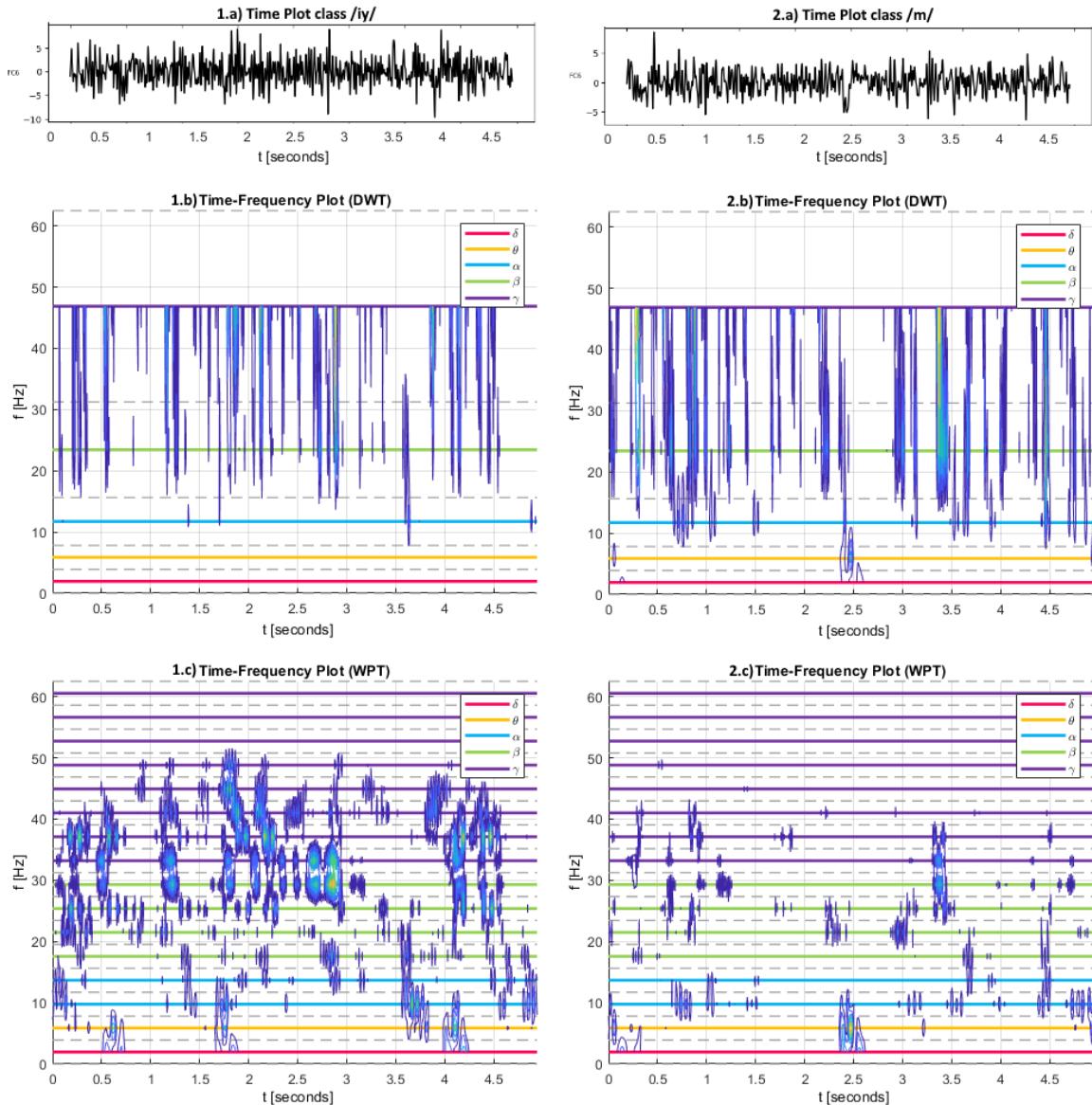


Figure 4.6: Examples of Wavelet energy analysis by sub-band.

Besides, Figure 4.6 shows examples of wavelet decomposition over FC6 channel signals from the same subject when imagined the class */iy/* (left column of plots) and */m/* (right column of plots). Plots 1.a) and 2.a) represent the signals in continuous time (seconds), while plots 1.b)-2.b) and 1.c)-2.c) show the time-frequency energy locations after applying DWT and WPT, repectively, over the signals with the MODWPT function.

Such time-frequency plots show the central frequencies of each filter used in the wavelet decompositions. These central frequencies are represented with horizontal lines, that are distinguished by colors associated with  $\delta$ ,  $\theta$ ,  $\alpha$ ,  $\beta$  and  $\gamma$  sub-bands. Also, the dashed gray lines represent the borderlines between filters.

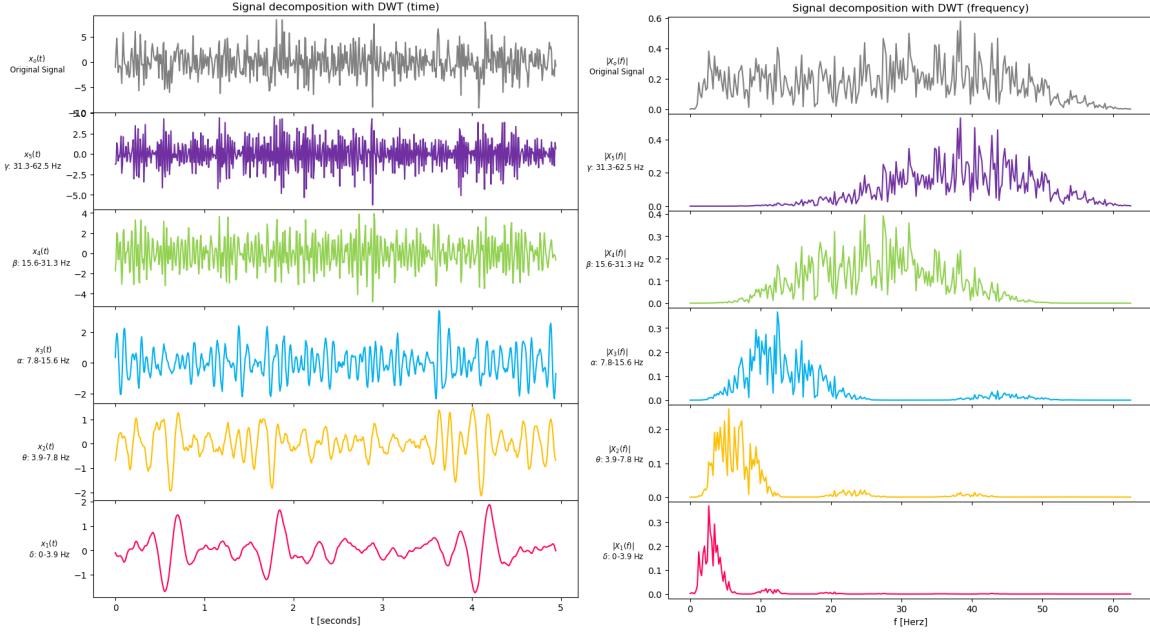


Figure 4.7: Signal decomposition with DWT.

It is noticeable that the concentration of energy vary among sub-bands and time intervals. However, in both cases of using DWT, the energy is more concentrated in  $\beta$  and  $\gamma$  sub-bands, even though the more detailed decompositions by using WPT show more dispersion of energy. For the class  $/iy/$ , such dispersion was more notable than for the class  $/m/$  by presenting energy concentrations varying from 2 to 50 Hz along the timesteps.

Due to that, each wavelet approach presents different decompositions of the signals, and as result, different feature vectors are built. In consequence, it was reasonable to compare the classification performance achieved with both wavelet approaches.

In order to remark their differences, Figures 4.7 and 4.8 show the signal decomposition applying DWT and WPT, respectively, of the same example from Figure 4.6 when the subject imagined the class  $/iy/$ .

The first plot columns show the signal in continuous time followed by each resulting wavelet coefficients set (distinguished by the same sub-band colors of Figures 4.5 and 4.6). While the second plot columns present their corresponding spectral components computed with the FFT. In the case of DWT spectral plots (Figure 4.7), it is obvious the band limits of each decomposition with just some small components overlapped in the borderlines of the bands. This effect is notable between  $\alpha$  and  $\beta$  sub-bands, as well as  $\beta$  and  $\gamma$  sub-bands.

On the other hand, it can be noticed in WPT spectral plots (Figure 4.8) some components present in sub-bands distant from the band limits of the corresponding decomposition. This phenomenon is notable in some  $\beta$  and  $\gamma$  sub-bands. (for example in the  $\beta$  sub-bands 19.5-23.4 Hz and 23.4-27.3 Hz, as well as in the  $\gamma$  sub-bands 50.8-54.7 Hz and 54.7-58.6 Hz) However, such components have less magnitude compared with those concentrated in the sub-band associated with the decomposition.

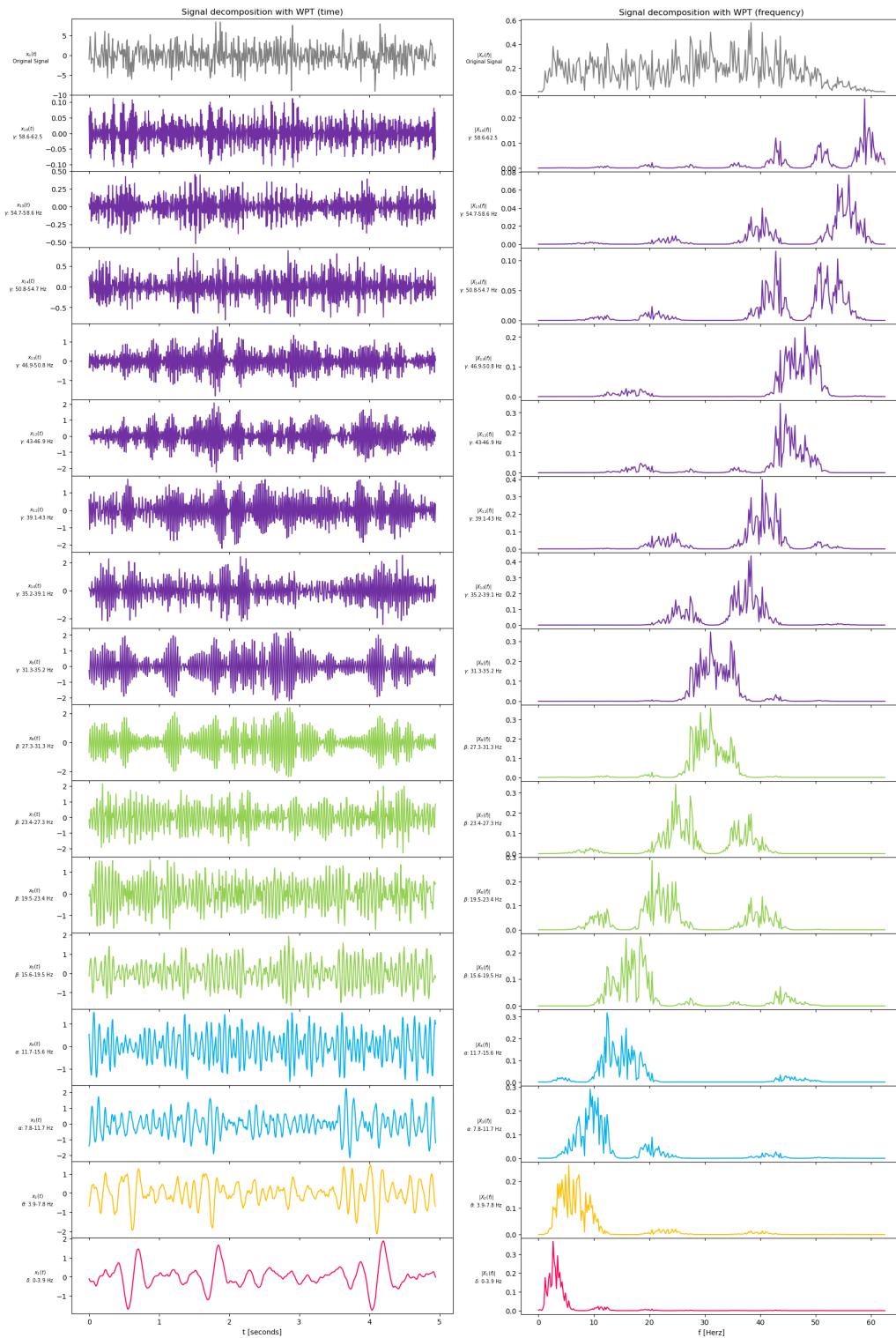


Figure 4.8: Signal decomposition with WPT.

## Feature extraction

Table 4.3: Features extracted (and their identifiers) based on [4] and proposed subset highlighted.

ID	Name	ID	Name	ID	Name
$F_1$	Mean	$F_2$	Mean(Abs)	$F_3$	Max
$F_4$	Max(Abs)	$F_5$	Min	$F_6$	Min(Abs)
$F_7$	Max+Min	$F_8$	Max-Min	$F_9$	EHF
$F_{10}$	Curvelength	$F_{11}$	Energy	$F_{12}$	Avgle
$F_{13}$	Spec entr.	$F_{14}$	6th power	$F_{15}$	Integral
$F_{16}$	Stand. Dev	$F_{17}$	Variance	$F_{18}$	Skewness
$F_{19}$	Kurtosis	$F_{20}$	Sum	$F_{21}$	Median

The same framing procedure, as in [1], was followed to extract features over the processed signals (trimmed signals in the case of overt speech and wavelet coefficients in the case of imagined speech). This procedure consisted of framing each signal by segments of approximately 10% of their lengths and with an overlap of 50% between each segment, resulting in 19 segments per signal.

Then, the features indicated in Table 4.3 were computed on every segment. These features were proposed in [4], and all of them were used in [1, 2]. Apart from computing those 21 features, a subset of three features were selected as a proposal to perform independent classification experiments and compare the performance achieved with both feature sets.

The three selected features are highlighted in Table 4.3 and they were selected because of their following characteristics defined in [4]:

- Curve length (CL): It is useful for observing amplitude and frequency changes and dimensionality of the signal. It is defined with the following equation:

$$CL_\ell = \sum_{m=0}^{M-1} |x[m-1] - x[m]| \quad (4.3.1)$$

Where  $M$  is the total number of points in the  $\ell$  segment and  $x[m]$  is the sample point value in  $m$ , considering 0 de first position in the segment and  $M - 1$  the last.

- Average Nonlinear Energy (ANE): It is a measure of energy proportional to both signal amplitude and frequency. This feature is defined as:

$$ANE_\ell = \frac{1}{M} \sum_{m=0}^{M-1} NE[m] \quad (4.3.2)$$

From where  $NE[m]$  is defined as:

$$NE[m] = \begin{cases} x^2[m] - x[m-1]x[m+1] & \text{if } 0 < m < M-1 \\ 0 & \text{otherwise} \end{cases} \quad (4.3.3)$$

In [4] it is specified that a Hanning window must be weighted with all the  $M$  values of  $NE$ . However, none window was applied over  $NE$  values for this work, like in [1, 2].

- Spectral Entropy (SE): This feature quantifies regularity and order in the signal with the following equation:

$$SE_\ell = \sum_{\kappa=0}^{M-1} P[\kappa] \log_2 \{P[\kappa]\} \quad (4.3.4)$$

Then,  $P[\kappa]$  is defined as:

$$P[\kappa] = \frac{|X[\kappa]|^2}{MF_s}, 0 \leq \kappa \leq M-1 \quad (4.3.5)$$

Where  $X$  is the Fourier transform of the signal  $x$  computed with the FFT of  $M$  points.

It is important to remark that the database also has the feature vectors extracted from the preprocessed signals (those after applying the BPF) in files named *all\_features\_noIca.mat* and the computation of these features was reproduced based on the definitions of [4]. This reproduction was made to implement the feature extraction step in a DSP for a course.

Nevertheless, some of the reproduced feature values differed from those of the database. For example, the spectral entropy and the curve length presented higher values in the database compared with the reproduced values. Besides, these and some other features presented higher values than the rest in the database, which was the motivation to applied a selected normalization in [16] for such features.

Due to the discrepancy and high values, the code provided by the University of Toronto was revised and compared with the definitions in [4]. After this revision, it was noticed that the features with such problems, which were stored in the database and used in [1, 2, 16], were wrongly implemented. For this work, the features were computed with the new implementation based on the definitions in [4].

As mentioned before, the framing procedure for this work was similar to that in [1]. In the case of overt speech, the resulting 19 sequential temporal sub-vectors, each formed by the 3 or 21 features extracted from each segment, represent a whole static feature vector to feed any classifier in this approach.

However, for imagined speech, each temporal sub-vector is formed by the 3 or 21 features from each wavelet coefficient set. This process is represented in Figure 4.9 with the 5 coefficient sets obtained after applying DWT. Each pair of dashed red or black vertical lines indicates the interval in which the  $L$  features of the  $M_\ell$  temporal sub-vector were extracted.

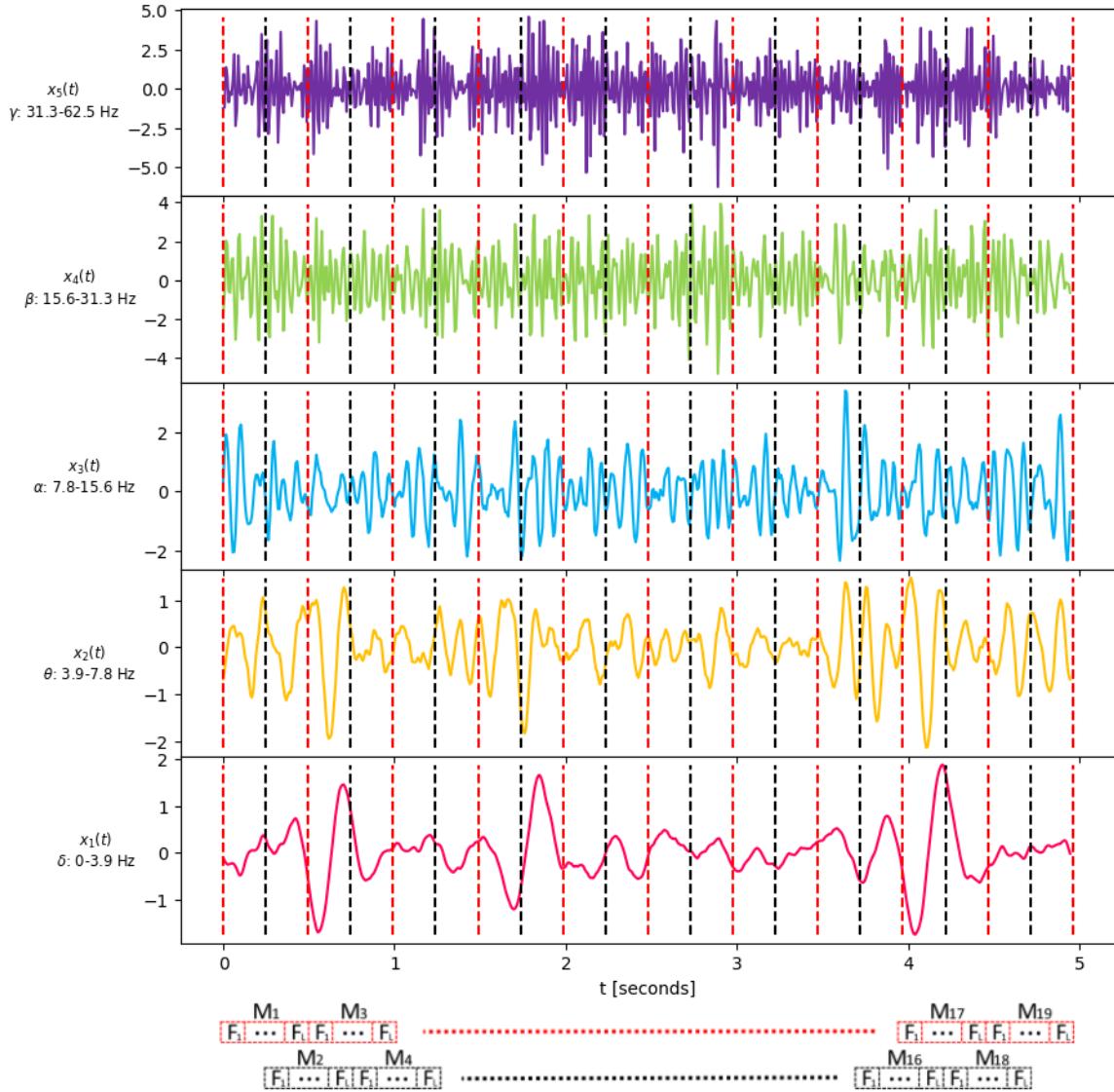


Figure 4.9: Signal framing from DWT decomposition.

Nonetheless, using temporal sub-vectors from all the coefficient sets would produce feature vectors of high dimensionality, which involves more computation in the classification step. For example, in the simplest case, 3 features and applying DWT, the feature vector dimension would be 285. Whereas, for the most complex case, 21 features and applying WPT, would be of 6384. This dimension variabilities would also require an architecture per case for the MLP since the number of hidden neurons depends on the input dimensions.

For this problem, a PCA was performed over each feature. This method reduced and fixed the dimensionality of feature vectors among both wavelet approaches (DWT and WPT). Figure 4.10 summarizes the processes involved in the PCA methodology applied in this work. The first step consisted of standardizing each feature across all the  $\omega$  sub-bands  $SB$  (i.e., by columns according to the green ovals in Figure 4.10). This step is crucial because PCA is sensible to feature scales.

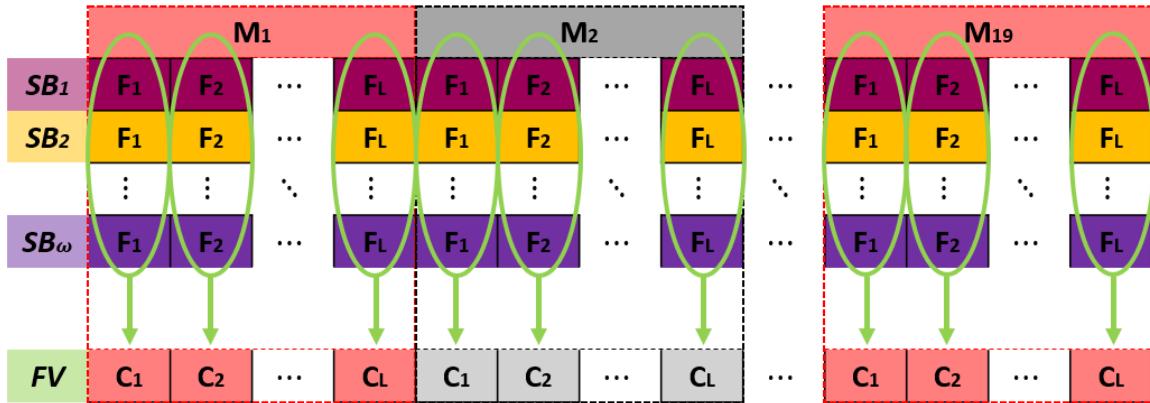


Figure 4.10: PCA over features by wavelet coefficient sub-band.

According to [25], if the data is not standardized, PCA might determine that the direction of maximal variance more closely correspond with the feature with highest scale values, which could not be necessarily true.

After standardization, the PCA components were computed across the  $SB$  sub-bands (again, this is indicated with the green ovals in Figure 4.10). Here, the  $L$  features from each segment represented the observations, while the sub-bands were the  $\omega$  variables. This process generated a set of components per feature (or observation), from which just the component with most variance was selected to build each temporal sub-vector of the resulting  $FV$  vector represented in Figure 4.10

The selected component's variances when using 3 and 21 features as observations are summarized in Table 4.4 and 4.5, respectively. The values in the tables represent the average  $\pm$  the variance computed across all the samples for the specified binary set, segment, and wavelet approach for  $FC6$  channel data.

It can be noticed that the selected components by using 3 features (Table 4.4) had, on average, variances above 97%, which means that they contained most of the information when data was reduced. On the other hand, the selected components by using 21 features (Table 4.5) conserved, on average, 94% of the information in the best case of DWT, which was reduced when using WPT. This means that PCA was more inconsistent for 21 features than for 3 features across all the binary sets, frames and wavelet approaches.

It is also important to mention that previously another approach was performed to reduce the size of feature vectors from imagined speech signals. This approach consisted of correlating the feature vector from each EEG signal with each feature vector from the resulting wavelet coefficient sets of the same signal. Then, based on the resulting ranking, just the most correlated coefficient sets were selected from all the decompositions.

In Figure 4.11 is represented this selection process with all the samples used in this work.  $\nu_O$  represent the feature vectors from the EEG signal and  $\nu_{SB}$  signify a specific wavelet coefficients set. Also,  $\rho$  is the Spearman correlation between each pair of vectors.

Table 4.4: Average Variance of selected components across wavelet approaches and segments using 3 features.

CV	DWT				WPT			
	-Nasal	+Nasal	-Bilabial	+Bilabial	-Nasal	+Nasal	-Bilabial	+Bilabial
1	98.48±2.83	98.04±3.34	98.23±3.04	98.47±3.03	99.37±1.28	99.17±1.53	99.26±1.36	99.35±1.41
2	98.1±3.56	98.24±2.94	98.28±3.11	97.95±3.7	99.21±1.54	99.26±1.34	99.29±1.31	99.12±1.69
3	97.97±3.49	98.27±2.74	98.28±2.95	97.75±3.65	99.14±1.56	99.29±1.13	99.29±1.24	99.03±1.66
4	98.18±3.34	98.29±3.09	98.32±3.23	98.05±3.28	99.25±1.42	99.29±1.34	99.32±1.35	99.19±1.45
5	98.39±3.04	98.58±2.61	98.56±2.85	98.31±2.96	99.34±1.33	99.36±1.43	99.36±1.42	99.31±1.27
6	98.44±3.03	98.59±2.63	98.48±2.99	98.51±2.68	99.37±1.26	99.34±1.45	99.33±1.46	99.41±1.09
7	98.59±2.76	98.69±2.71	98.71±2.74	98.49±2.74	99.42±1.21	99.45±1.16	99.46±1.18	99.39±1.22
8	98.49±3.14	98.69±2.45	98.62±3.02	98.47±2.69	99.36±1.44	99.48±0.97	99.41±1.35	99.39±1.17
9	98.49±3.07	98.85±2.25	98.76±2.7	98.39±2.93	99.36±1.43	99.52±0.95	99.47±1.29	99.34±1.25
10	98.61±2.84	98.78±2.41	98.86±2.56	98.35±2.86	99.41±1.29	99.48±1.14	99.51±1.23	99.31±1.25
11	98.59±2.52	98.81±2.39	98.87±2.21	98.35±2.85	99.43±1.01	99.49±1.08	99.54±0.91	99.31±1.22
12	98.57±2.92	98.85±2.28	98.77±2.54	98.51±2.95	99.39±1.36	99.52±1	99.47±1.21	99.38±1.29
13	98.51±3.17	98.63±2.53	98.53±2.92	98.58±3.01	99.34±1.56	99.43±1.1	99.36±1.44	99.39±1.37
14	98.38±3.24	98.67±2.48	98.46±3.04	98.53±2.88	99.34±1.39	99.46±0.99	99.37±1.27	99.41±1.23
15	98.61±2.8	98.57±2.76	98.54±2.8	98.69±2.76	99.42±1.24	99.39±1.29	99.41±1.17	99.42±1.39
16	98.73±2.69	98.69±2.72	98.69±2.75	98.77±2.62	99.46±1.22	99.44±1.27	99.45±1.22	99.47±1.26
17	98.81±2.52	98.68±2.69	98.75±2.48	98.76±2.77	99.51±1.11	99.47±1.05	99.48±1.09	99.51±1.08
18	99±2.09	98.77±2.44	98.86±2.46	99.02±1.77	99.6±0.84	99.49±1.12	99.53±1.09	99.61±0.66
19	98.83±2.01	98.75±2.45	98.69±2.43	98.96±1.7	99.54±0.74	99.47±1.14	99.47±1.04	99.59±0.63

Table 4.5: Average Variance of selected components across wavelet approaches and segments using 21 features.

CV	DWT				WPT			
	-Nasal	+Nasal	-Bilabial	+Bilabial	-Nasal	+Nasal	-Bilabial	+Bilabial
1	80.92±16.55	81.59±14	98.37±5.75	98.17±6.16	79.17±8.75	79.65±9.37	74.88±10.23	80.4±11.28
2	81.69±15.62	82.01±13.54	98.33±6.05	97.71±7.24	79.46±8.71	79.21±9.16	73.36±8.9	79.04±10.59
3	81.52±14.83	82.17±13.66	98.39±6.22	97.47±7.64	79.60±8.76	79.01±8.37	73.09±9.19	78.68±10.78
4	80.66±15.93	82.82±13.75	96.99±8.26	97.13±8.14	79.76±8.69	79.60±8.97	73.02±9.41	78.54±10.72
5	80.5±15.39	82.48±13.49	96.72±8.33	96.98±8.72	79.75±8.51	79.89±9.21	72.78±9.39	77.94±10.74
6	81.09±14.69	80.73±14.45	96.29±9.33	96.68±8.82	79.71±8.55	79.87±9.38	72.65±9.09	77.73±10.82
7	80.69±14.57	80.62±14.09	95.97±10.49	96.36±10.04	80.5±8.85	79.69±9.92	72.82±9.05	77.44±10.76
8	81.32±14.43	80.72±13.99	96.02±10.09	96.36±9.17	80.32±9.14	80.12±9.17	72.88±9.19	78.18±10.43
9	81.55±14.56	81.25±13.52	95.89±10.06	96.76±8.77	79.98±9.21	80.34±8.85	72.42±9.15	78.36±10.24
10	80.64±14.49	81.47±13.2	95.69±10.34	97.05±7.79	79.95±9.66	80.39±8.76	72.51±9.14	77.89±10.61
11	80.39±14.46	80.55±14.09	95.59±10.39	96.32±9.32	80.72±9.27	80.69±8.73	72.49±9.13	77.72±10.76
12	80.53±14.14	80.36±13.56	95.51±10.64	95.85±10.32	80.89±8.93	80.2±9.13	72.21±0.09	77.42±10.87
13	80.88±14.59	81.03±13	95.52±10.45	95.75±10.58	80.61±9.08	79.71±9.39	72.32±9.09	77.23±10.81
14	81.05±13.89	80.55±14.05	95.48±10.85	95.92±9.89	80.66±9.13	79.95±9.23	72.25±8.99	76.91±10.69
15	80.86±13.71	79.75±14.61	95.01±11.18	96.2±10.51	80.91±9.05	80.19±9.39	72.23±8.89	76.59±10.68
16	80.76±14.55	79.43±14.62	95.15±11.21	94.54±11.56	80.75±9.17	80.41±9.48	72.46±8.79	76.91±10.75
17	80.14±14.78	79.02±14.89	94.44±12.33	94.07±12.09	80.86±8.79	80.76±9.36	72.34±8.94	76.89±10.47
18	80.09±14.93	80.41±14.43	94.09±12.85	94.65±11.89	80.66±9.05	80.37±9.02	72.45±8.78	76.66±10.18
19	79.22±14.5	80.33±14.63	95.01±11.25	96.26±9.14	80.2±8.93	80.05±8.73	73.61±10.25	77.98±10.98

Besides, each straight colored curve represents all the absolute correlation coefficient values sorted in ascending order as a result of correlating all the  $\nu_{SB}$  vectors (associated with sub-bands distinguished in the legend) with their corresponding  $\nu_O$  vector.

Also, the horizontal dashed red line is a threshold fixed in 0.7, and the vertical dashed lines indicate the points where each corresponding correlation curve crosses the threshold.

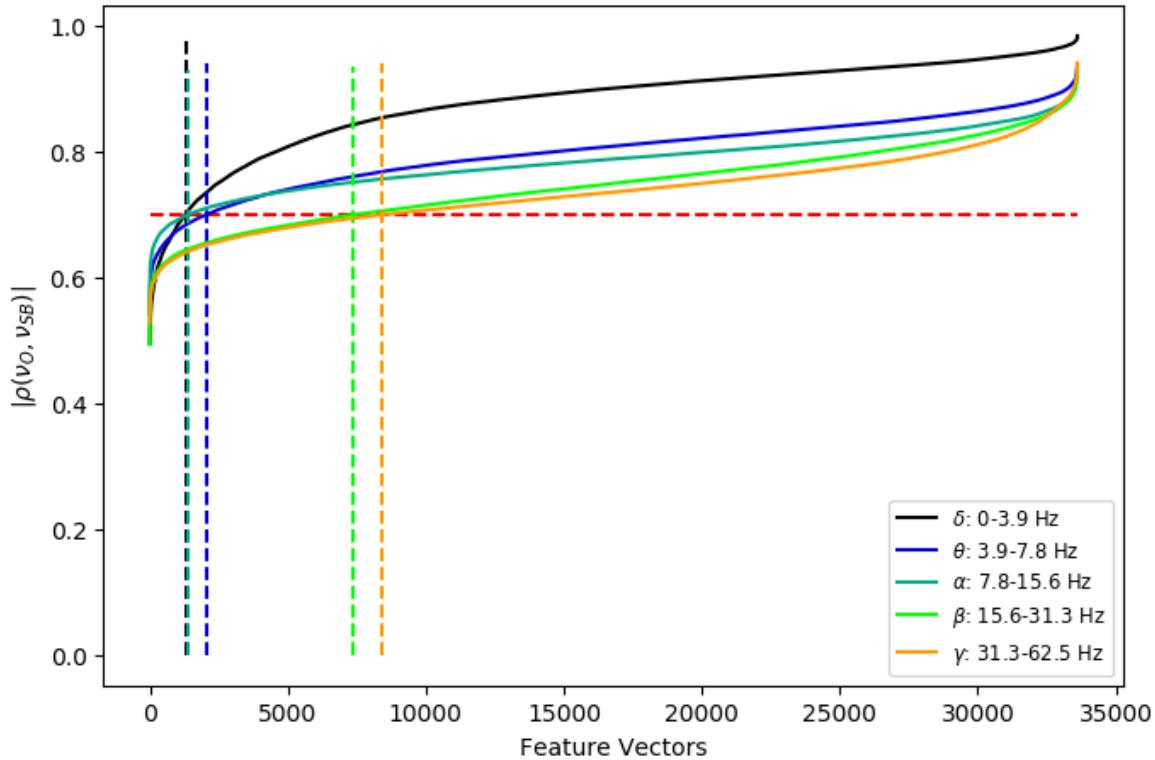


Figure 4.11: Spearman correlation using DWT and 3 features.

Just the three  $\nu_{SB}$  vectors with more amount of correlation coefficients above the threshold were selected for the classification step.

In Figure 4.11, where DWT and 3 features were computed, the three  $\nu_{SB}$  vectors correspond to  $\theta$ ,  $\delta$ , and  $\alpha$  sub-bands. Indeed, for the other cases (21 features and DWT, 3 or 21 features and WPT) the corresponding sub-bands of the selected vectors were the same. In Table 4.6 are shown, for all the cases (wavelet approach and computed features), the ranking, the associated sub-bands with the wavelet coefficient sets, their corresponding range in Hz, and the percentage of correlation coefficients (from a total of 33604) above the threshold per case. Also, the three selected feature vectors are highlighted in Table 4.6, which were selected due to their notorious percentage of correlation coefficients comparing with the rest of the feature vectors.

This process was done with Kendall and Pearson correlation too. However, the results obtained with Kendall correlation were the worst, from which in most of the cases the best sub-band correlated obtained 76.67% of correlation coefficients above the threshold.

While for Pearson correlation, apart of obtaining similar results as with Kendall correlation, was necessary to carry the Kolmogorov-Smirnoff test to decide if the correlation values were significant, i.e. if  $\nu_{SB}$  and  $\nu_O$  came from the same continuous distribution ( $H_0$ ) or not ( $H_a$ ). The Kolmogorov-Smirnoff test was carried with a 5% significance level, which rejected  $H_0$  for all  $\nu_{SB}$  and  $\nu_O$  pairs.

Table 4.6: Spearman correlation rankings per case and selected sub-bands

	Rank	3 features			21 features		
		Subband	Range	Percentage	Subband	Range	Percentage
DWT	1	$\delta$	0-3.9	96.14	$\delta$	0-3.9	44.93
	2	$\alpha$	7.8-15.6	95.91	$\theta$	3.9-7.8	34.5
	3	$\theta$	3.9-7.8	93.92	$\alpha$	7.8-15.6	28.07
	4	$\beta$	15.6-31.3	78.07	$\beta$	15.6-31.3	19.85
	5	$\gamma$	31.3-62.5	75.03	$\gamma$	31.3-62.5	17.48
WPT	1	$\delta$	0-3.9	96.39	$\delta$	0-3.9	96.14
	2	$\theta$	3.9-7.8	95.38	$\alpha$	7.8-11.7	95.91
	3	$\alpha$	7.8-11.7	90.29	$\theta$	3.9-7.8	93.92
	4	$\gamma$	31.3-35.2	80.63	$\alpha$	11.7-15.6	78.07
	5	$\beta$	27.3-31.3	78.9	$\beta$	15.6-19.5	75.03
	6	$\gamma$	35.2-39.1	78.35	$\beta$	19.5-23.4	74.56
	7	$\gamma$	39.1-43	74.94	$\beta$	27.3-31.3	69.29
	8	$\beta$	23.4-27.3	68.02	$\gamma$	31.3-35.2	68.25
	9	$\alpha$	11.7-15.6	67.77	$\beta$	23.4-27.3	61.86
	10	$\beta$	15.6-19.5	66.21	$\gamma$	35.2-39.1	55.21
	11	$\beta$	19.5-23.4	62.95	$\gamma$	39.1-43	47.51
	12	$\gamma$	43-46.9	62.18	$\gamma$	43-46.9	35.31
	13	$\gamma$	46.9-50.8	54.79	$\gamma$	46.9-50.8	25.43
	14	$\gamma$	50.8-54.7	48.89	$\gamma$	58.6-6	7.45
	15	$\gamma$	58.6-6	46.54	$\gamma$	50.8-54.7	6.52
	16	$\gamma$	54.7-58.6	46.28	$\gamma$	54.7-58.6	2.76

For the reasons exposed above, Kendall and Pearson correlations were discarded. However, although the best percentages were obtained with Spearman correlation, this overall processing step was discarded since the classification using these feature vectors provided poor score values (around 56% of accuracy on average). These results were obtained with some tests varying MLP and SVM parameters. Nevertheless, based on the scores, it seemed that the classifiers recognized almost always just one class with these feature vectors.

## Classification

Once the feature vectors were built, which were composed of 19 temporal sub-vectors with 3 or 21 features (or components) each, the general architecture for the classification consisted of establishing one classifier per existing EEG channel (for this work 62). Figure 4.12 depicts this architecture, where each classifier received as input feature vectors, of 19  $M$ -temporal sub-vectors with  $L$  features each, from one particular channel. Besides, the classifiers used in this work were the MLP (ANN) and SVM.

As a first step, each classifier received just the corresponding training feature vectors to update their parameters (this is the learning step). Therefore, a validation methodology (which is explained further) was followed to train the 62 classifiers and compute some scores per channel with all the training data. These scores were obtained to register the performance achieved with each classifier. The score types computed were: accuracy, recall, precision, and F1.

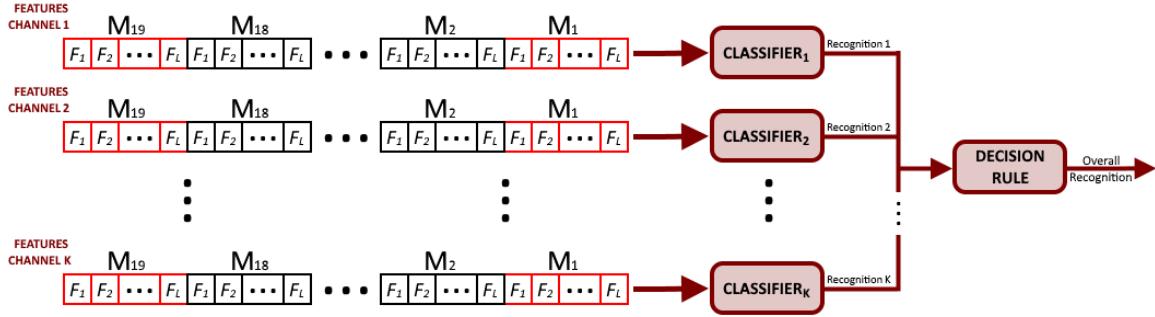


Figure 4.12: Architecture for Vector-based classifiers.

---

**Algorithm 1** Channel's selection

---

**Input:**

$M$ : Score matrix

$n$ : Number of channels to select

**Output:**

$\nu$ : List of selected channels

```

1: Initialize:  $S \leftarrow \mathbf{0}$ ,  $C \leftarrow \mathbf{0}$ , list  $\nu$ 
2: for  $i = 1$  to  $u$  do
3:    $S_i \leftarrow \text{sort}(M_i)$ 
4:    $C_i \leftarrow \text{argsort}(M_i)$ 
5: end for
6: Flatten  $S$  and  $C$ 
7:  $\vec{c} \leftarrow \text{sortscores}(S, C)$ 
8:  $i = 1$ 
9: repeat
10:  if  $\vec{c}_i$  not in  $\nu$  then
11:    Add  $\vec{c}_i$  to list  $\nu$ 
12:  end if
13:   $i \leftarrow i + 1$ 
14: until  $\dim(\nu) = n$ 

```

---

In the particular case of the MLP, this learning step was performed 20 times with different initializations since this classifier is sensitive to the initial parameter's values. After storing the scores from the 20 initializations, the averages of each score type across all the iterations were computed and stored per channel to register their performance.

Next, the second step (named score improvement in Figure 4.1) consisted of selecting just a channel's subset based on the best scores obtained in the previous step. For this step, just one score type was selected as a reference.

The Algorithm 1 illustrates this process, in which the input matrix  $M$  contained the scores and had dimensions of  $u$  by  $n$  (number of subjects and channels, respectively). Besides, the output list  $\nu$  stored the  $n$  channel's identifiers from which were obtained the highest scores.  $M$  is a matrix because each row represents the scores obtained across all the channels when all data except subject's data associated with such row were used to train the classifier.

Table 4.7: Selected channels in [1, 2], and those selected in this work for overt and imagined speech using MLP and SVM.

Ranking	Kara One		Overt speech				Imagined speech			
	Channel	$\rho$	MLP		SVM		MLP		SVM	
			Channel	F1	Channel	F1	Channel	F1	Channel	F1
1	FC6	0.3781	C3	61.9474	FZ	82.9932	FC6	53.9541	AF4	83.0296
2	FT8	0.3758	FC6	60.5664	C2	82.0485	CP5	53.5614	OZ	82.0796
3	C5	0.3728	PO3	60.3989	CP6	81.7204	F5	51.1207	P8	82.0088
4	CP3	0.372	F4	60.377	TP7	76.1885	FT7	50.6732	PZ	82.0088
5	P3	0.3696	C1	60.0264	CP3	74.2852	C6	50.549	F1	82.0088
6	T7	0.3686	CP3	59.8149	O1	73.9607	FT8	50.4072	AF3	82.0088
7	CP5	0.3685	CPZ	59.6918	P3	73.2558	C5	50.2671	POZ	81.8386
8	C3	0.3659	F5	58.9745	FC4	73.1938	C1	50.1141	FC1	81.6886
9	CP1	0.3626	FC5	58.9437	CB2	71.7073	FC4	49.8666	CB2	81.2775
10	C4	0.3623	C5	58.6208	P2	71.6779	F8	49.6962	C3	81.2775

Also,  $S$  and  $C$  represent two matrices (of  $u$  by  $n$ ) that store, by row and in descending order, the sorted score values and their channel's identifiers, respectively. After the storage, both matrices were flattened to sort again  $C$  based on the values of  $S$ . Then, the resulting values were stored in  $\vec{c}$ . Finally, the first  $n$  unique values of  $\vec{c}$  were added to the output list  $\nu$ .

For this work, the value of  $n$  was set to 10. This decision was made to perform experiments comparable with [1, 2]. Additionally, some classification experiments were carried, in which the number of selected channels was varied. After finishing these experiments, the best scores were obtained by using 10 channels.

On the other hand, the score type used to select a subset of channels was F1. This decision was made again after performing some experiments using the accuracy and F1 scores in the process described in Algorithm 1. The selected channels based on the F1 score provided better performance in the third step than those based on the accuracy score. Besides, it was noticed that the classifiers from the selected channels based on the accuracy classified almost all samples as one particular class.

In Table 4.7 are shown the selected channels and the scores obtained with MLP and SVM classifiers for overt and imagined speech (no wavelet decomposition performed). These channels were also the result of using 3 features and  $\pm$ Bilabial binary sets. Additionally, the 10 selected channels in [1, 2] are shown in the column Kara One in Table 4.7. Notice that the score used there was the Pearson correlation (denoted with the letter  $\rho$ ) between feature vectors from acoustic, and EEG imagined speech signals.

Therefore, in Table 4.7 can be noticed that the resulting selected channels vary among classifiers, mental states, and experimental approaches. Figure 4.13 illustrates these differences, where it can be seen that for overt speech samples, just one channel (CP3) was common among all the cases. Whereas, for imagined speech samples was not any channel in common. Besides, the best scores were obtained by using SVMs; however, comparing both, Table 4.7 and Figure 4.13, these selected channels were not common with MLP, neither with [1, 2].

To complement the explanation of Algorithm 1, in Table 4.8 is shown the channel

selection process of the same case of Table 4.7 by using an SVM for overt speech. The Scores and Channels represent, respectively, the  $S$  and  $C$  matrices from the algorithm (once finished the for cycle). Then, the orange row contains the selected channels (those in yellow in  $S$  and  $C$  matrices), which were selected in the order exposed in the Ranking row.

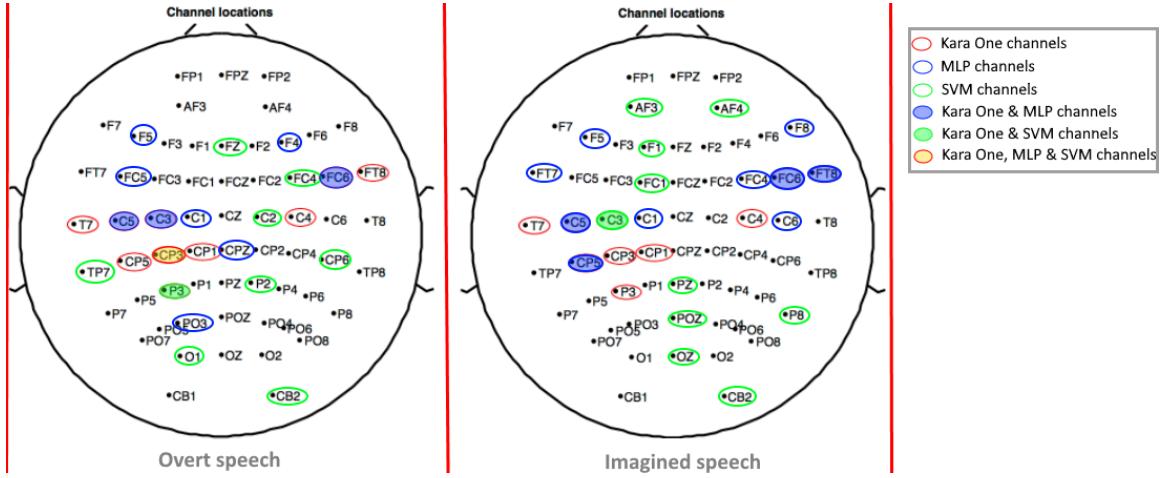


Figure 4.13: Selected channels in [1, 2], and those selected in this work for overt and imagined speech using MLP and SVM (Original Figure taken from [3] for its adaptation in this work).

Table 4.8: Example of channel's selection based on the best (average) F1 scores.

Ranking	1	2	3	4	5	6	7	8	9	10
Scores	81.7204	72.7026	71.7073	71.5593	70.2795	69.5274	69.2439	68.8333	67.5215	67.1659
	75.6084	73.8739	73.1938	72.2644	71.6779	71.6777	71.6774	69.482	69.1112	68.6422
	73.7852	73.1618	70.4938	70.4938	70.013	69.1816	67.9257	66.5594	65.625	65.2672
	79.4051	77.1167	74.3707	70.8944	69.8651	69.0848	68.0152	67.7833	67.3944	66.9954
	82.9932	82.0485	76.1885	74.2852	73.9607	73.2558	72.2323	70.8371	70.5964	69.6508
Channels	CP6	FZ	CB2	PO4	FC1	PO7	T8	CZ	CPZ	F6
	TP7	CP6	FC4	FZ	P2	C2	FC5	F1	FC6	T8
	O1	C2	FC4	OZ	FCZ	CP4	P5	P4	PO7	FT7
	CP6	FZ	TP7	CPZ	C5	FC2	FC6	F1	FC4	O1
	FZ	C2	TP7	CP3	O1	P3	FC4	FC6	FC5	F1
Selection	FZ	C2	CP6	TP7	CP3	O1	P3	FC4	CB2	P2

Notice that in this process, some elements in matrix  $S$  had high F1 scores; however, they were discarded because the associated channels had been previously appended to the selection list (i.e. such channels had been identified before with a higher F1 score). These cases are highlighted in both matrices with pale yellow color.

It is also important to mention that other criteria were proposed previously to perform the channel's selection. Nevertheless, as happened with the other decisions, after performing some experiments, the best scores in the third step were obtained with the criterion exposed in Algorithm 1. The generalities of these other proposed criteria are described as follows:

- Average Scores: This criterion consisted of computing the average scores from  $S$  between repeated channels based on  $C$  items. Then, the average values are sorted and stored in  $\vec{c}$ , following the rest of the process as in Algorithm 1.

- Weighted Scores: This criterion was similar to Average Scores. However, instead of averaging the repeated channels, they were summed and divided by  $u$ .
- Accuracy with F1 Scores: This criterion consisted of counting the occurrences of each channel from the accuracy and F1 score matrices  $C$ . Then, the occurrence's lists are sorted in descending order to form the selected channel's ranking by the following priorities: 1) It was in both matrices, 2) it occurred more times in the F1 matrix than in the accuracy matrix, 3) it occurred in just F1 matrix but more than one time.

Finally, as a third step for Vector-based classification, just the  $n$  classifiers that received feature vectors from the selected channels were used. However, in this step, both training and test data were introduced to the new architecture composed of  $n$  classifiers, which each output a recognized class label. Once these  $n$  labels were obtained, the decision rule (also shown in Figure 4.12) was based on the class label's count.

If 70% (or more) of the classifiers predicted the correct class label associated with the input sample, then such class is stated as the overall recognition. Otherwise, another class was randomly stated. Indeed, as binary classification experiments were performed for this work, the sample is associated with the other binary class if less than 70% of the classifiers predicted the correct class label.

Once collected the overall recognitions from all the training and test samples, all the score types were performed (accuracy, recall, precision, F1). However, just the accuracy score was reported in the classification results for simplification. Also, for the particular case of the MLP, this third step was executed 20 times again, and the average accuracies  $\pm$  the standard deviations are reported.

### 4.3.2 Spatio-Temporal

#### Encoding

As a first step for the NeuCube, input signals were required to be encoded in spike trains since the neurons of the SNN captured, with their biological neuron model, the temporal information from them, while the connections represent the spatial information. Therefore, the problem was to select an encoding method that ensured the preservation of relevant information from the EEG signals by compressing them in spike trains.

Due to this concern, in [26] were tested different temporal spike encoding methods with several signal types that model a wide variety of behaviors. The approach adopted in that work was to compare the encoders' performance by computing some metrics based on the error between the original and reconstructed (or decoded) signals. The reconstruction of the signal refers to the process of transforming the resulting spike trains to EEG signals.

Hence, the error metrics computed in [26] were the SNR, RMSE, and R-squared measure. Thus, based on the resulting metric values from several performed experiments, it was concluded in [26] that the SF encoding method outperformed the others used there, which

were BSA, TBR, and MW.

Notice that carrying experiments across encoding methods and error metrics (i.e., similarly as [26]) to select an adequate encoder for overt and imagined speech samples was out of scope for this work. Nevertheless, two encoding methods were selected and mutually compared based on their classification performance: SF [27] and BSA [28].

Additionally, instead of testing with several signals, the focus in this work consisted of performing optimization steps in each encoding method to minimize the RMSE.

SF was used due to the results and conclusions provided in [26]. Hence, the optimization process consisted of performing a grid search of the additive threshold value that provided the least RMSE. An optimal threshold value was obtained from and applied to each EEG signal. Therefore, the threshold values used for the grid search were the range from the EEG signal amplitudes by a factor varying from 0.001 to 0.8 with a step size of 0.001.

On the other hand, BSA was selected because it had been successfully applied in EEG signals in other earlier works [29, 30]. Moreover, in [26] was not performed any optimization procedure, and consequently, it was not explored this encoder thoroughly. In this work, a DEA [31] with constraints was used to obtain, similarly as for SF grid search, the optimal encoding per EEG signal by minimizing the RMSE.

For BSA optimization, the weighting factor, the crossover probability, and the number of generations were set to 0.02, 0.7, and 100, respectively. Additionally, as this encoding is based on an FIR filter, the constraints for the DEA were the filter order, the bandpass (radians per sample), and the BSA threshold with values between [3, 50], [0.1, 1], and [0.2 \*  $maxValue$ , 2 \*  $maxValue$ ] (where  $maxValue$  is the maximal value of the normalized signal), respectively.

In Table 4.9 and 4.10 are shown for overt and imagined speech, respectively, the optimal values, as well as the RMSE, obtained after applying the DEA for BSA encoding of the 12 EEG signals from the *FC6* channel corresponding to the class */iy/* from the subject *S<sub>1</sub>*. Also, the last row shows the mean of each parameter value  $\pm$  their variance.

Notice that although these signals belong to the same particular case (same subject, class, and channel), the optimal parameter's set founded vary per signal. Therefore, these results confirm the importance (and consequently the computational cost) of customizing the encoding process per signal.

Finally, Figure 4.14 illustrates an example of encoding EEG signals with both methods: SF and BSA. The plots from the first column belong to an overt speech signal, while those from the second column belong to an imagined speech signal. Indeed, both signals represent the same case: subject *S<sub>1</sub>*, class */iy/*, channel *FC6*, trial 1.

Thus, in both columns, the first plot shows the original EEG signal in black, as well as the reconstructed signal after applying the BSA and SF encoding in dashed blue and green lines, respectively. Notice that for these examples, SF reconstructed the overt speech signal

Table 4.9: Optimal values obtained after applying the DEA for BSA encoding: Overt Speech.

Samples	Order	Bandpass	Threshold	RMSE
1	28.271	0.17199	1.5127	0.038272
2	38.921	0.11952	1.8157	0.035735
3	30.141	0.11233	1.666	0.045256
4	36.762	0.10304	1.7186	0.023799
5	30.952	0.14482	1.65	0.043209
6	38.093	0.18224	1.5182	0.059114
7	31.103	0.25417	1.0692	0.097987
8	14.593	0.18057	1.6381	0.058363
9	19.213	0.15574	1.5316	0.052366
10	26.39	0.13468	1.7515	0.018461
11	25.488	0.12689	1.4348	0.051495
12	21.431	0.11995	1.88	0.046047
<b>Mean</b>	<b>28.45±7.55</b>	<b>0.151±0.04</b>	<b>1.5989±0.21</b>	<b>0.0475±0.02</b>

Table 4.10: Optimal values obtained after applying the DEA for BSA encoding: Imagined Speech.

Samples	Order	Bandpass	Threshold	RMSE
1	28.85	0.22533	1.2691	0.092413
2	30.896	0.13097	1.4116	0.13636
3	45.052	0.19549	1.3476	0.15121
4	48.274	0.20386	1.1463	0.14691
5	31.041	0.22677	1.0035	0.15148
6	10.811	0.1405	1.2954	0.16745
7	12.54	0.29571	1.3338	0.091356
8	13.618	0.17725	1.4635	0.12126
9	20.785	0.25031	1.3084	0.1089
10	31.744	0.16863	1.3277	0.11905
11	8.1956	0.28021	1.3053	0.11948
12	14.674	0.13437	1.4901	0.11043
<b>Mean</b>	<b>24.71±13.4</b>	<b>0.203±0.06</b>	<b>1.3085±0.13</b>	<b>0.1264±0.02</b>

better than BSA, while for the imagined speech signal was the opposite case.

Besides, the second pair of plots show the spike trains after applying BSA, which provides just positive spikes. Whereas, the third pair of plots correspond to the encoded signal after applying SF, which provides positive and negative spikes.

## Classification

After all the samples were encoded, the following methodology was performed to select the data for a grid search that would allow optimizing the parameters of the NeuCube for

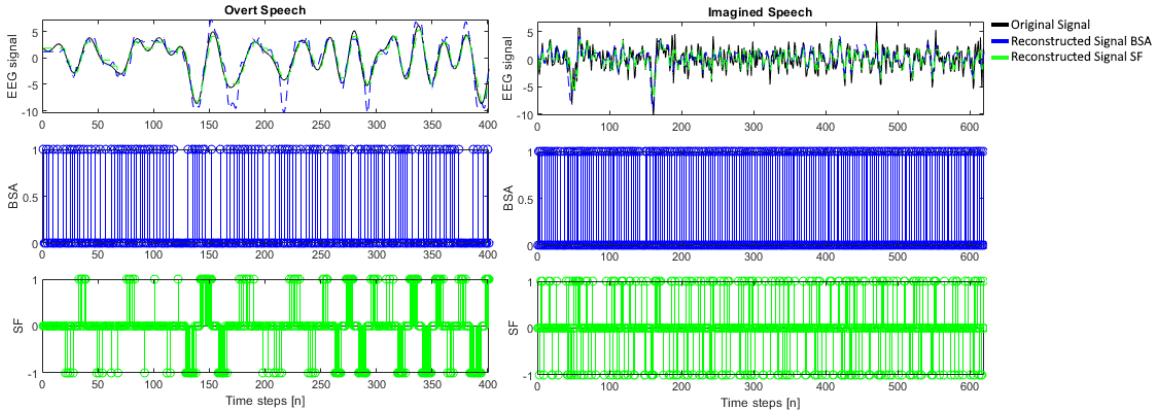


Figure 4.14: Encoded signal examples using SF and BSA.

the classification step:

1. Overt and imagined speech data were selected from the binary classes ( $\pm$ Nasal or  $\pm$ Bilabial), and the codification applied (BSA or SF). The combination of such cases resulted in 4 data sets per mental task (overt or imagined speech).
2. Then, for each data set, just the samples of  $S_1$  were taken as test data since that subject produced more trials. Experiments by taking the samples of the other subjects as test data were not performed in this step due to the long-time they would consume following this methodology.
3. Next, several classification experiments were performed by varying the percent of excitatory connections ( $r_+$ ) and LIF threshold ( $\vartheta$ ) once the training and test data were stated per data set. Therefore, the overall accuracy (sum of training and test accuracies divided by two) was registered per case, and the same initialization was used across all the cases (i.e., just one experiment was performed per case).
4. Later, the parameter pairs (8 in total) that provided the best overall accuracy per mental task and data set were selected for the next step in this methodology. Table 4.11 shows the overall accuracies across all the cases, and the best, and consequently selected, are highlighted.
5. As the last step, these best parameter pairs were used again (with their corresponding mental task and data set samples) but now following the validation methodology described in the next chapter, which provides accuracies that represent when each subject's data were used for testing. This step was done to select the subject's test data for the grid search. Table 4.12 shows the resulting overall accuracies from 10 different initializations. Besides, the best results for overt and imagined speech samples are highlighted, which were also the data set samples used for the grid search of such mental tasks.

Finally, based on the selected data sets ( $\pm$ Nasal, SF,  $S_5$  test data for overt speech;  $\pm$ Bilabial, SF,  $S_4$  test data for imagined speech), Table 4.13 shows the values used for some NeuCube parameters in a grid search per mental task. Performing the grid search for all

Table 4.11: Overall accuracies for NeuCube data set parameter's selection.

		Overt Speech		Imagined Speech	
		BSA	SF	BSA	SF
Nasal	0.7	0.1	58.35	59.14	57.87
	0.7	0.05	58.61	58.97	47.22
	0.5	0.1	59.62	61.64	59.53
	0.5	0.05	53.53	55.42	48.01
	0.7	0.1	59.67	58.26	50.37
Bilabial	0.7	0.05	53.00	57.69	46.03
	0.5	0.1	62.08	60.02	52.21
	0.5	0.05	58.48	58.83	49.59
					55.33

Table 4.12: Overall accuracies for NeuCube samples selection for grid search.

		Overt Speech		Imagined Speech	
	Subject	BSA	SF	BSA	SF
Nasal	$S_1$	59.68	55.86	56.25	54.87
	$S_2$	58.23	56.44	54.53	54.95
	$S_3$	55.69	57.55	54.62	55.22
	$S_4$	58.71	58.22	54.61	55.08
	$S_5$	59.68	60.74	51.34	55.55
Bilabial	$S_1$	56.93	58.24	52.18	57.96
	$S_2$	58.66	56.01	54.46	56.10
	$S_3$	53.01	57.52	55.51	57.17
	$S_4$	58.66	56.75	53.83	58.31
	$S_5$	57.27	57.62	54.71	54.94

Table 4.13: Grid search values used for the NeuCube.

Parameters	Values
$r_+$	0.3, 0.5, 0.7
$\vartheta$	0.05, 0.07, 0.08, 0.1, 0.11, 0.13
$\pm A$	0.001, 0.005, 0.01
$\pm D$	0.001, 0.005, 0.01

the NeuCube parameters was not realizable due to time-consuming.

After these grid searches, the parameter set founded was stated for the definitive classification of all samples. This definitive classification consisted of computing the training and test scores with the validation methodology using 20 different NeuCube initializations but with the same parameter set.

### 4.3.3 Mixed Approach

The SSN performs the classification step in a Spatio-temporal approach by multiplying all EEG channel data at once with their corresponding input weights per time step. However, it is considered as a mixed approach classifier in this work since it was fed with EEG preprocessed signals (following the corresponding Spatio-temporal steps) as well as with feature vectors (which implied processing and feature extraction steps of the Vector-based approach). Despite the differences between step approaches, the experiments were performed with the same conditions for both input data types.

Contrary to the NeuCube experiments, for the SNN were used two different neuron models: LIF and Izhikevich. This last was not used in the NeuCube because the variables model's combinations can represent 27 different behaviors of the neuron, which was not viable to test each of them per case. However, it was possible to apply Izhikevich model for the SSN due to the DEA of [31] employed during the training step. Additionally, performing the DAE over Izhikevich model had the advantage of using just the behaviors that prevailed on every new generation instead of testing all of them.

Furthermore, by using the DEA, each individual was a vector composed of the 62 input weight values on the interval  $[-0.01, 0.01]$  and the neuron model parameters to tune. In the case of the LIF model, these parameters were the refractory period and the threshold ( $\vartheta$ ) values on the intervals  $[0, 6] \in \mathbb{Z}$  and  $[0.1, 1]$ , respectively. On the other hand, the only parameter used for the Izhikevich model was the behavior's identifier, from which its interval values were  $[0, 27] \in \mathbb{Z}$ . Additionally, the population size per generation was set to 10 times the length of the individuals; i.e., the number of individuals for the LIF and the Izhikevich model was 640 and 630, respectively.

Notice that the SSN training step (explained in Chapter 3 of this thesis) implied an internal optimization process, which consisted of minimizing the error by selecting the best individuals for the next generation in the DEA. Consequently, the SSN did not require any other score improvement as in the Vector-based approach (with a channel's selection step) or the NeuCube (with a grid search).

Figure 4.15 shows the number of errors obtained per generation for both mental tasks (overt and imagined speech) during the training step, from which the LIF neuron model,  $\pm$ Bilabial sets, and  $S_3$  test data were used. However, in these examples, 21-feature vectors were used as input for overt speech; while, EEG signals were used for imagined speech.

Hence, the blue and red curves represent, respectively, the best and the worst-case per generation. Besides, the green curve represents the average number of errors across all the individuals per generation, and the yellow lines are the  $\pm$ standard deviation from the average error, which had few variations across the generations.

Notice that for both cases in the examples of Figure 4.15, after 200 generations, the blue curve started to decreased less. In the end, the SSN configured with the best individual (from the 500th generation) misclassified 137 and 157 samples for overt and imagined speech, respectively, during the training step.

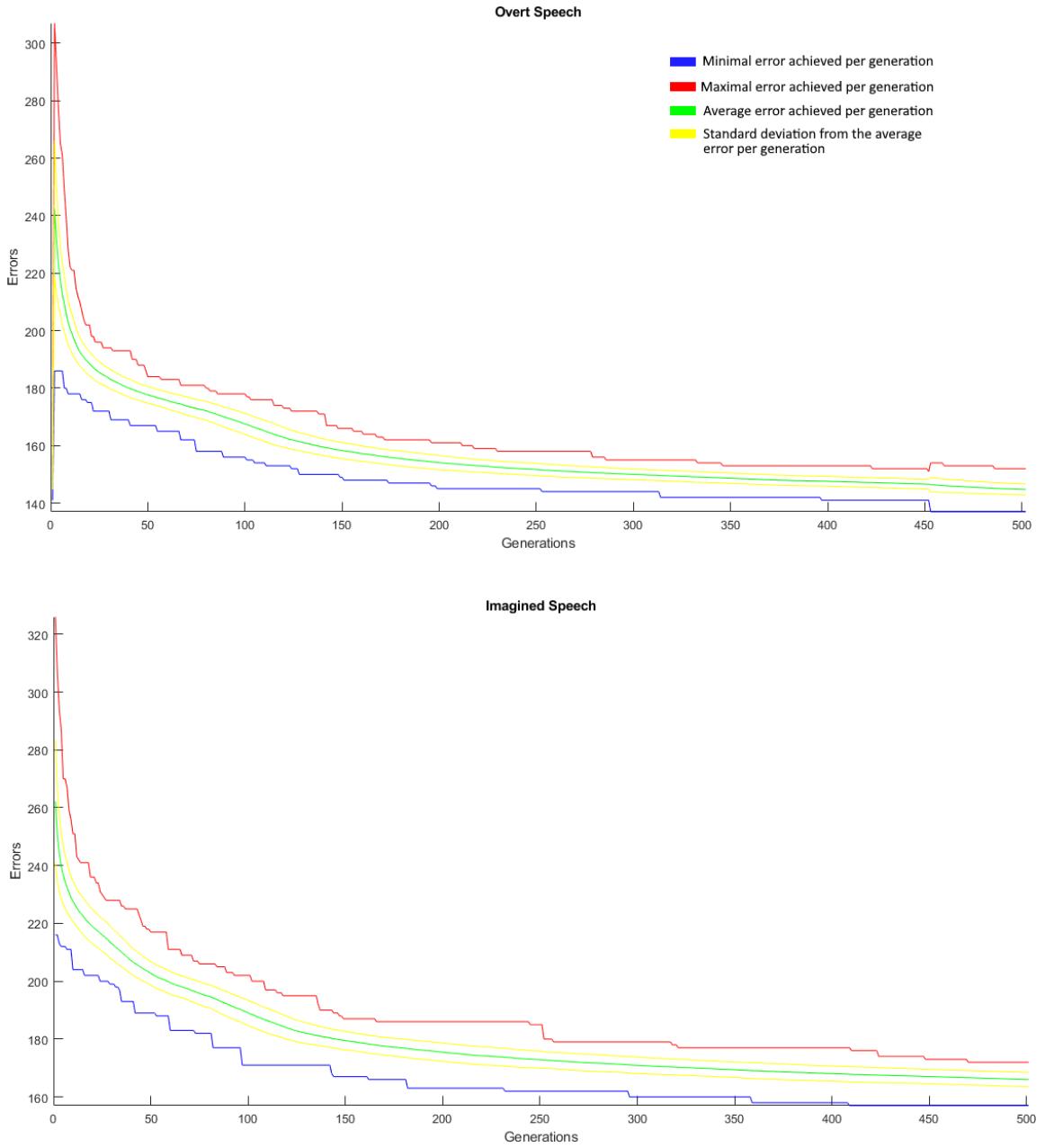


Figure 4.15: Number of errors per generation during SSN training for overt and imagined speech examples.

These last and best SSN configurations were used for the test step, which consisted of passing the test samples through the SSN. Then, their firing rates were computed and compared with the AFRs from both classes to associate them with the closest. Finally, the scores were computed based on the errors produced in the classification.

# EXPERIMENTAL FRAMEWORK

---

In this chapter is described the validation methodology used with all the classifiers from Vector-based, Spatio-temporal, and Mixed Approaches. Furthermore, it is explained how the data were split and the reasons to do that for the classifier's performance evaluation. Besides, the selected parameters used in all the experiments for each classifier are shown at the end of this chapter.

## 5.1 Validation Methodology

Based on what is mentioned in [32], Figure 5.1 shows, inside red squares, the data used by the classifiers in each step of the validation methodology followed in this work. As can be seen, the first step consisted of holding the data of one subject ( $S_m$ ) out from the rest to use it later for testing, while the remaining data would be used to train the classifier. This process was done to perform subject-independent experiments similar to [12, 13, 1, 2]. The reason for performing the experiments in this way was to test if the classifiers could recognize overt and imagined speech from any subject's samples.

Then, in the second step, the training data is used to compute scores that would provide an idea of the classifier recognition's capability. These scores were obtained with a k-fold cross-validation process to provide unbiased results. This argument is valid since, on every round of the process, the test fold data were unseen by the time the classifier was trained with the rest data.

Therefore, the training scores reported in the next chapter of this work refers to the case when test fold data were used as input to the classifier. Also, the same parameter's initialization was set in all rounds to perform the classifier's training with the same conditions. Besides, a stratified version was used to balance the sample's percent of each class present on the training and test data every round. According to what is mentioned in [33], the main advantage of using stratified k-fold cross-validation is the experimental variance reduction, which makes it easier to identify the best of the methods under consideration.

Next, the third step consisted of training again the classifier but with all training data. This step was done to build a final version of the classifier. Finally, the fourth step consisted of testing the trained classifier with the subject data held out at the beginning. Thus, the test scores reported in the next chapter of this work resulted from this final step.

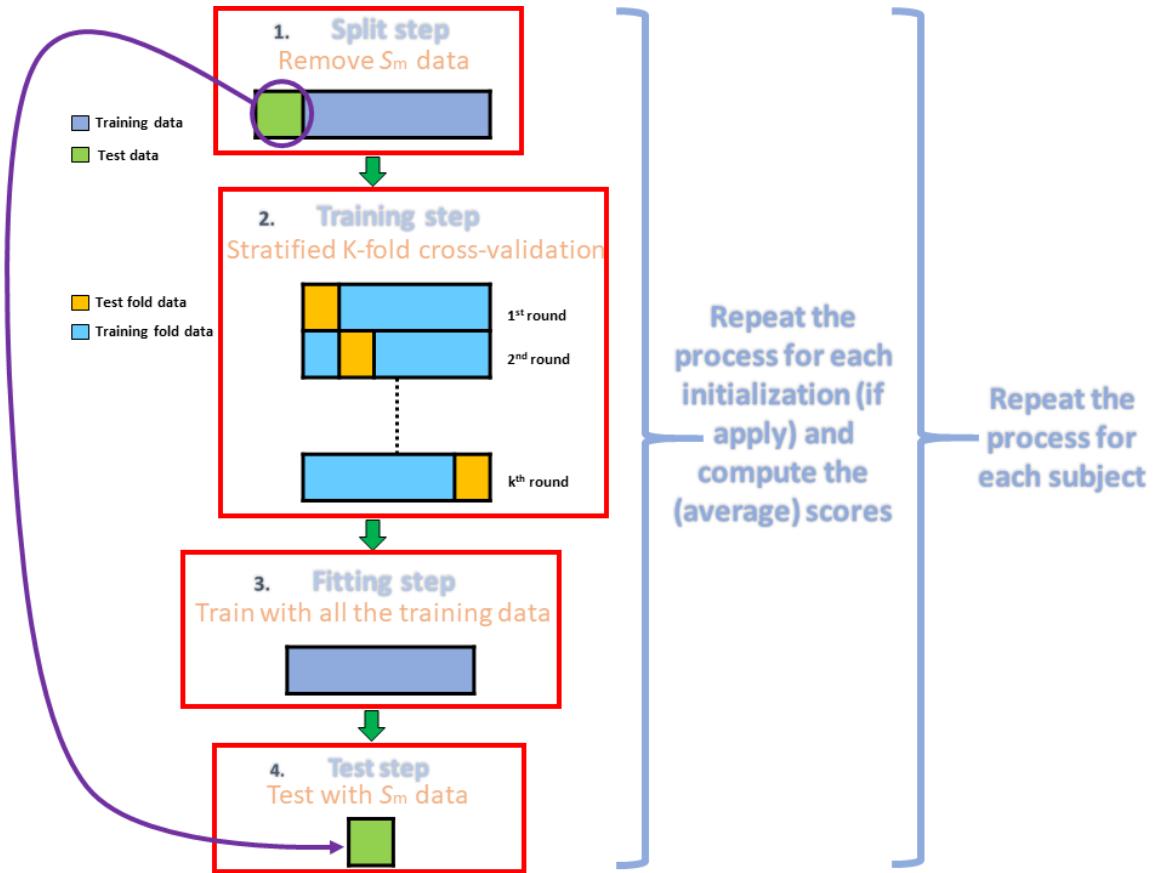


Figure 5.1: Data used on each step of the validation methodology followed.

These four steps were repeated several times in case the classifier was sensitive to initial parameters, and the score's averages  $\pm$  their standard deviation across the iterations were computed at the end. Besides, the data in step 2 were shuffled to produce different data order on every iteration. Additionally, these steps were repeated by removing data from a different subject in step 1 per time. Thus, the scores reported in this work are identified by whom were the data test.

To complement the explanation of this methodology, Algorithm 2 shows the details of the followed steps. The matrix  $X$  contained the data points of each sample by channel. Then, for this work,  $\varsigma$  was conformed by the accuracy, recall, precision, and F1 score names. Besides, the number of subjects  $u$  was 5, the number of experiments  $n$  (which would cause different classifier's initializations) was set to 20 (if applied), and the number of folds was set to 5. Additionally, the output matrices  $S_f$  and  $S_t$  contained, respectively, the scores obtained by using training (test fold) and test data, and from which each row represents which subject data was held out from  $X$ .

Also, the *for* cycles in steps 2 and 6 represent, respectively, the subject's data selection and the different classifier's parameters initializations per case (second and first bracket in Figure 5.1). Table 5.1 shows the relation between Figure 5.1 and Algorithm 2 steps.

---

**Algorithm 2** Validation methodology

---

**Input:**

$X$ : Data samples  
 $\varsigma$ : List of score's names to compute  
 $u$ : Number of subjects  
 $n$ : Number of experiments to perform  
 $f$ : Number of folds in the training data

**Output:**

$S_f$ : Training scores (training fold data)  
 $S_t$ : Test scores (by subject data)

```
1: Initialize: matrix  $S_f$ , matrix  $S_t$ 
2: for  $i = 1$  to  $u$  do
3:   Initialize: list  $\varsigma_f$ , list  $\varsigma_t$ 
4:    $X_f \leftarrow X \setminus \{\text{data of subject } i\}$ 
5:    $X_t \leftarrow \{\text{data of subject } i\} \in X$ 
6:   for  $j = 1$  to  $n$  do
7:     Initialize: classifier's parameters, list  $\gamma_f$ , list  $\gamma_t$ 
8:     Shuffle  $X_f$  data
9:     for  $k = 1$  to  $f$  do
10:     $T \leftarrow X_f \setminus \{k \text{ fold data}\}$ 
11:     $E \leftarrow \{k \text{ fold data}\} \in X_c$ 
12:    Train the classifier with the initial parameters and  $T$ 
13:    Test the classifier with  $E$ , add recognitions to  $\gamma_f$ 
14:  end for
15:  Train the classifier with the initial values and  $X_f$ 
16:  Test the classifier with  $X_t$ , add recognitions to  $\gamma_t$ 
17:  Compute  $\varsigma$  scores from  $\gamma_f$  and  $\gamma_t$  recognition's lists, then add them to  $\varsigma_f$  and  $\varsigma_t$ , respectively
18: end for
19: Compute mean  $\varsigma$  scores from  $\varsigma_f$  and  $\varsigma_t$ , then add them to  $S_f$  and  $S_t$ , respectively
20: end for
```

---

Table 5.1: Relationship of steps between Figure 5.1 and Algorithm 2.

Steps Figure 5.1	Steps Algorithm 2
1	3 and 4
2	9 to 14
3	15
4	16

Notice that, based on Algorithm 2, the test fold data recognitions were stored in  $\gamma_f$  and the scores are computed after the *for* cycle's ending (steps 9 to 14), instead of computing the scores per fold to compute the score averages later. This criterion was chosen due to the conclusions from [33]. In [33] is mentioned that this criterion avoids biased results, especially under high-class imbalance and using F1 score, which is the case in this work.

## 5.2 Parameters' Selection

Tables 5.2, 5.3, and 5.5 show, respectively, the most important parameters used for the Vector-based approach, NeuCube and SSN classifiers. This means that the scores reported in the next chapter were obtained by using these configurations on each classifier.

The *scikit* learn library, written in Python language, was used to implement the two Vector-based classifiers from this work: MLP and SVM. Also, their parameters (shown in Table 5.2) were selected after performing some experiments, which consisted of varying some parameter values and comparing the accuracy obtained with each parameter set following the validation methodology described in the previous section. It is important to note that these experiments were performed by using just some data sets due to the time would consume using all the data configurations.

In the case of MLP, these varied parameters were the number of iterations (or epochs) with values of 50, 100, and, 200; as well as the hidden neuron's activation function with the hyperbolic tangent (*tanh*) and the rectified linear unit (*relu*) function. As shown in Table 5.2, the configuration that provided the best scores in these experiments was composed of a *tanh* activation function and 50 iterations.

The experiments that used 50 iterations outperformed those with 100 or 200 iterations because, according to the developer site, the *lbfgs* weight optimization algorithm converges faster with small databases, which is the case in this work. Besides, these configurations were tested with normalized and non-normalized input data, and the best results were obtained by using normalized data.

Furthermore, just the kernel type was varied for the SVM with a second-grade polynomial and a radial basis function (*rbf*). As shown in Table 5.2, the configuration that provided the best scores was composed of an *rbf* kernel. Thus, similarly as with MLP, the SVM parameters were tested with normalized and non-normalized data, from which the best results were obtained by using non-normalized data.

On the other hand, the NeuCube's parameters, shown in Table 5.3, were selected after the grid search was executed for overt and imagined speech samples. In the previous chapter was explained that these grid searches were done over the intervals shown in Table 4.13. Nonetheless, the grid searches were done with two different parameter sets, which each had the same parameter intervals from Table 4.13 but with their own  $\vartheta$  values: **1.** 0.07, 0.1, 0.13, and **2.** 0.05, 0.08, 0.11. Besides, each grid search was performed twice with different initializations to ensure that the selected parameters were the best for these experiments.

Table 5.2: Selected parameters for Vector-based classifiers.

<b>Multilayer Perceptron (MLP)</b>	
Number of hidden neurons (just one layer)	<i>300</i>
Activation function of hidden neurons	<i>tanh</i>
Weight optimization	<i>lbfgs</i>
Regularization term parameter	$1 \times 10^{-4}$
Maximum number of iterations	<i>50</i>
<b>Support Vector Machine (SVM)</b>	
Kernel type	<i>rbf</i>
kernel coefficient	<i>scale</i>
Tolerance for stopping criterion	$1 \times 10^{-3}$

Table 5.3: Selected parameters for the Neucube.

<b>NeuCube</b>
<b>SWC</b>
Radius for candidate connections: <i>2.5</i>
Weight values interval: $\pm 0.1$
$r_+$ : <i>0.3</i>
<b>LIF</b>
$\vartheta$ : <i>0.13</i> , Refractory period: <i>6</i>
<b>STDP</b>
Interspike intervals: $\pm 10$
$+A$ : <i>0.001</i> , $-A$ : <i>0.005</i>
Synaptic weight values interval: $\pm 2$
<b>deSNN</b>
$+D$ : <i>0.005</i> , $-D$ : <i>0.001</i> , $\alpha$ : <i>1</i> , <i>KNN</i> : <i>5</i>
Modulation factor: <i>0.8</i>

Table 5.4 summarizes the parameters for each grid search performed that provided the best overall accuracy (i.e., the sum of training and test accuracies divided by two). In the cases when overt speech samples were used, the resulting parameters were the same with two different initializations by using the parameter set 1 (highlighted in Table 5.4).

However, the grid searches with different initializations but with the same parameter's intervals varied when imagined speech samples were used. For this reason, classification experiments for imagined speech samples were performed with the two different parameter sets highlighted in Table 5.4 using the  $\pm$ Bilabial sample's organization.

Then, it was selected the parameter's set that provided the best accuracies, which were the same as for overt speech. Due to that, in Table 5.3 is shown just one parameter set that was used for both: overt and imagined speech.

Table 5.4: NeuCube parameter values that provided the best overall accuracy and organized by set used.

Samples	Set	Initialization	$\vartheta$	$+A$	$-A$	$+D$	$-D$	$r_+$	Accuracy
Overt	1	1	0.13	0.001	0.005	0.005	0.001	0.3	67.15%
		2	0.13	0.001	0.005	0.005	0.001	0.3	67.79%
	2	1	0.05	0.005	0.01	0.001	0.005	0.7	66.88%
		2	0.05	0.001	0.01	0.001	0.01	0.3	68.17%
Imagined	1	1	0.13	0.001	0.001	0.01	0.005	0.3	64.29%
		2	0.07	0.001	0.005	0.001	0.001	0.7	66.46%
	2	1	0.08	0.001	0.01	0.001	0.005	0.3	64.39%
		2	0.05	0.001	0.001	0.001	0.01	0.7	64.29%

Table 5.5: Selected parameters for the SSN.

Single Spiking Neuron (SSN)
<b>LIF</b>
$\vartheta$ : [0.1, 1.0]
Refractory period: $[0, 6] \in \mathbb{Z}$
<b>Izhikevich</b>
27 behaviours varying:
$a, b, c, d, I$
<b>Optimization</b>
Number of generations: 500
Crossover probability: 0.7
Weighting factor: 0.1

As it has been mentioned in the previous chapter, the SSN has an internal optimization process with a DEA, which tries to find the best parameter configurations based on the constraints stated beforehand. Thus, in Table 5.5 are shown the parameters used for the DEA, as well as the interval parameters from both neuron models used with the SSN classifier: LIF and Izhikevich.

Notice that the single parameter in Izhikevich model represents the behavior's identifier, which was composed by particular values of the variables  $a, b, c, d$ , and  $I$ . These particular values and the behavior's names that the neuron simulates are shown in Table 5.6. Thus, the names and values of the behaviors 0-19 were obtained from [34] and [35], respectively. While, the names and values of the behaviors 20-26 came from [36] and [37], respectively.

As mentioned in the previous chapter, once the 500 generations of the DEA were performed, the best individual (optimal parameter's values) was selected for the test data. Due to that, in Tables 5.7, 5.8, 5.9, and 5.10 are shown, respectively, the optimal parameter's values obtained with overt speech, imagined speech, DWT, and WPT samples. Besides, as mentioned before, each subject row represents when such subject was held out from the rest data to test the classifier.

Table 5.6: Izhikevich behaviors used in SSN DEA.

ID	Name	Values: $a, b, c, d, I$
0	Tonic spiking	0.02, 0.2, -65, 6, 14
1	phasic spiking	0.02, 0.25, -65, 6, 0.5
2	Tonic bursting	0.02, 0.2, -50, 2, 15
3	Phasic bursting	0.02, 0.25, -55, 0.05, 0.6
4	Mixed mode	0.02, 0.2, -55, 4, 10
5	Spike frequency adaptation	0.01, 0.2, -65, 8, 30
6	Class 1 excitible	0.02, -0.1, -55, 6, 0
7	Class 2 excitible	0.2, 0.26, -65, 0, 1.0
8	Spike latency	0.02, 0.2, -65, 6, 7
9	Subthreshold oscillations	0.05, 0.26, -60, 0, 0
10	Resonator	0.1, 0.26, -60, -1, 0
11	Integrator	0.02, -0.1, -55, 6, 0
12	Rebound spike	0.03, 0.25, -60, 4, 0
13	Rebound burst	0.03, 0.25, -52, 0, 0
14	Threshold variability	0.03, 0.25, -60, 4, 0
15	Bistability	1, 1.5, -60, 0, -65
16	depolarizing after-potential	1, 0.2, -60, -21, 0
17	Accomodation	0.02, 1.0, -55, 4, 0
18	Inhibition-induced spiking	-0.02, -1.0, -60, 8, 80
19	Inhibition-induced bursting	-0.026, -1.0, -45, 0, 80
20	Regular spiking neurons (excitatory)	0.02, 0.2, -65, 8, 10
21	Intrinsically bursting (excitatory)	0.02, 0.2, -55, 4, 10
22	Chattering (excitatory)	0.02, 0.2, -50, 2, 10
23	Fast spiking (inhibitory)	0.1, 0.2, -65, 2, 10
24	Low-threshold spiking (inhibitory)	0.02, 0.25, -65, 2, 10
25	Inhibitory neurons	[0.02, 0.1], [0.2, 0.25], -65, 2.0, [-6.0, 6.0]
26	Excitatory neurons	0.02, 0.2, [-65, -49] $\in \mathbb{Z}$ , [2, 9], [-15.0, 15.0]

These parameter's values were used in the classification experiments, which results are shown in the next chapter. Besides, the subscripts 1, 2, and 3 in the variables represent, respectively, when 3 features, 21 features, and EEG signal's data points were used as inputs.

Notice that in Tables 5.9, and 5.10, subscript 3 is not present, which means that SSN has not been implemented to deal with the additional dimension of these data that represents the wavelet decompositions. Besides, curiously the behavior obtained in all wavelet cases was 13.

Table 5.7: Optimal SSN parameters obtained from overt speech experiments.

		LIF						Izhikevich		
		Threshold			Refractory Period			Behavior		
Subjects		$\vartheta_1$	$\vartheta_2$	$\vartheta_3$	$\varrho_1$	$\varrho_2$	$\varrho_3$	$v_1$	$v_2$	$v_3$
$\pm$ Nasal	$S_1$	-0.16	0.11	0.03	1	4	5	19	19	19
	$S_2$	-0.13	0.66	1.05	1	2	4	15	20	4
	$S_3$	0.02	0.35	0.00	1	0	3	15	19	9
	$S_4$	-0.05	0.10	0.01	1	5	5	15	18	19
	$S_5$	-0.09	0.29	0.00	1	5	6	15	20	15
$\pm$ Bilabial	$S_1$	1.00	0.87	0.75	0	0	0	23	5	0
	$S_2$	0.81	1.08	0.93	0	1	4	23	5	1
	$S_3$	0.69	0.97	1.06	6	0	0	23	5	27
	$S_4$	0.74	0.85	1.00	8	0	0	12	5	1
	$S_5$	0.83	1.01	1.06	6	2	0	27	5	5

Table 5.8: Optimal SSN parameters obtained from imagined speech experiments.

		LIF						Izhikevich		
		Threshold			Refractory Period			Behavior		
Subjects		$\vartheta_1$	$\vartheta_2$	$\vartheta_3$	$\varrho_1$	$\varrho_2$	$\varrho_3$	$v_1$	$v_2$	$v_3$
$\pm$ Nasal	$S_1$	-0.18	0.13	0.05	4	0	0	19	0	6
	$S_2$	0.07	0.03	0.62	1	5	5	19	27	6
	$S_3$	0.02	0.35	0.00	1	0	3	15	19	9
	$S_4$	0.24	0.75	0.16	0	5	0	5	19	20
	$S_5$	0.25	0.03	1.02	3	0	4	5	0	0
$\pm$ Bilabial	$S_1$	0.71	1.04	0.07	1	0	6	15	6	15
	$S_2$	0.81	0.99	0.04	0	0	6	15	6	15
	$S_3$	1.10	0.76	0.03	0	0	6	15	5	15
	$S_4$	1.18	0.70	0.05	1	0	6	15	6	15
	$S_5$	1.15	0.98	0.01	0	0	6	15	5	15

Table 5.9: Optimal SSN parameters obtained from imagined speech experiments using DWT.

		LIF				Izhikevich	
		Threshold		Refractory Period		Behavior	
Subjects		$\vartheta_1$	$\vartheta_2$	$\varrho_1$	$\varrho_2$	$v_1$	$v_2$
$\pm$ Nasal	$S_1$	0.02	0.06	2	6	13	13
	$S_2$	0.03	0.37	4	2	13	13
	$S_3$	0.08	0.07	7	3	13	13
	$S_4$	0.02	0.06	3	0	13	13
	$S_5$	0.06	0.06	5	2	13	13
	$S_1$	0.02	-0.02	4	4	13	13
	$S_2$	0.05	0.06	5	1	13	13
	$S_3$	0.03	-0.02	5	4	13	13
	$S_4$	0.04	-0.02	6	4	13	13
	$S_5$	0.05	-0.03	6	0	13	13
$\pm$ Bilabial	$S_1$	0.02	-0.03	4	3	13	13
	$S_2$	0.03	0.08	2	3	13	13
	$S_3$	0.05	-0.02	5	4	13	13
	$S_4$	0.06	0.06	5	4	13	13
	$S_5$	0.06	0.05	5	2	13	13
	$S_1$	0.04	-0.03	6	4	13	13
	$S_2$	0.06	0.07	6	4	13	13
	$S_3$	0.03	0.06	0	4	13	13
	$S_4$	0.05	0.07	6	5	13	13
	$S_5$	0.04	0.09	6	1	13	13

Table 5.10: Optimal SSN parameters obtained from imagined speech experiments using WPT.

		LIF				Izhikevich	
		Threshold		Refractory Period		Behavior	
Subjects		$\vartheta_1$	$\vartheta_2$	$\varrho_1$	$\varrho_2$	$v_1$	$v_2$
$\pm$ Nasal	$S_1$	0.02	-0.03	4	3	13	13
	$S_2$	0.03	0.08	2	3	13	13
	$S_3$	0.05	-0.02	5	4	13	13
	$S_4$	0.06	0.06	5	4	13	13
	$S_5$	0.06	0.05	5	2	13	13
	$S_1$	0.04	-0.03	6	4	13	13
	$S_2$	0.06	0.07	6	4	13	13
	$S_3$	0.03	0.06	0	4	13	13
	$S_4$	0.05	0.07	6	5	13	13
	$S_5$	0.04	0.09	6	1	13	13
$\pm$ Bilabial	$S_1$	0.02	-0.03	4	3	13	13
	$S_2$	0.03	0.08	2	3	13	13
	$S_3$	0.05	-0.02	5	4	13	13
	$S_4$	0.06	0.06	5	4	13	13
	$S_5$	0.06	0.05	5	2	13	13
	$S_1$	0.04	-0.03	6	4	13	13
	$S_2$	0.06	0.07	6	4	13	13
	$S_3$	0.03	0.06	0	4	13	13
	$S_4$	0.05	0.07	6	5	13	13
	$S_5$	0.04	0.09	6	1	13	13

# RESULTS AND DISCUSSION

---

In this chapter are shown the results from all the classification experiments performed in this work using overt and imagined speech samples. The results are shown in tables that, for simplicity, contain the overall accuracies obtained in each case. These overall accuracies represent the sum of the training and test accuracies divided by two. In the particular cases of MLP and NeuCube, these accuracies came from the averages of the experiments with different initializations.

Furthermore, the experiments were conducted with the validation methodology described in the previous chapter. Due to this methodology, each row in the tables represents when a specific subject's data was held out from the rest to test the classifier (subject-independent experiments approach). Additionally, these results were obtained by the classifiers configured with the parameter's values also shown in the previous chapter.

Moreover, the experiments consisted of selecting the two less unbalanced data sets:  $\pm$ Nasal and  $\pm$  Bilabial with their corresponding processing steps for the Vector-based classifiers (MLP and SVM), Spatio-Temporal classifier (NeuCube), and the Mixed approach classifier (representing it as  $\text{SSN}_L$  and  $\text{SSN}_I$  when LIF and Izhikevich neuron models, respectively, were used). In each section, the results from each experiment set are analyzed, and the best scores, per case, are highlighted in the tables.

## 6.1 Vector-Based

Table 6.1 shows the overall accuracies obtained with overt speech samples using 3 or 21 features for the input vectors. For both cases, the best scores were obtained with the  $\pm$ Bilabial data set using the SSN with a LIF neuron model ( $\text{SSN}_L$ ). Notice that for all the classifiers and features used, the  $\pm$ Bilabial data set provided better results than  $\pm$ Nasal.

Furthermore, comparing the classifier's performances across the experiments shown in Table 6.1, the best were obtained with the  $\text{SSN}_L$ , followed by the  $\text{SSN}_I$ , SVM, and at last the MLP. Besides, according to these results, the use of 21 features provided slightly better results than using 3 features in the case of  $\text{SSN}_L$  for rows  $S_3$ ,  $S_4$ , and  $S_5$ .

Naturally, building vectors of 21 features implied more processing time consumption, but the SSN extracted more temporal information with 21 features than using just 3. Of course, more in-depth analysis in the future is required to select an optimal feature set to obtain both advantages: time consumption and performance.

Table 6.1: Overall accuracies overt speech (vector-based approach).

		3 FEATURES				21 FEATURES			
		MLP	SVM	SSN <sub>L</sub>	SSN <sub>I</sub>	MLP	SVM	SSN <sub>L</sub>	SSN <sub>I</sub>
$\pm$ Nasal	$S_1$	60.85	61.88	68.13	63.79	51.59	63.02	74.49	66.2
	$S_2$	59.73	60.72	67.94	70.32	52.24	62.64	70	67.33
	$S_3$	59.08	58.97	70.25	62.67	52.81	62.68	70.22	66.07
	$S_4$	59.01	63.71	71.79	56.37	53.39	62.86	68.89	57.68
	$S_5$	59.88	57.42	65.99	66.01	53.66	61.19	71.6	66.23
$\pm$ Bilabial	$S_1$	66.35	68.84	75.36	71.25	52.87	62.84	74.66	72.95
	$S_2$	67.69	66.14	72.77	71.595	55.47	64.32	71.99	71.63
	$S_3$	67.95	68.56	75.42	74.67	53.6	62.95	77.01	75.15
	$S_4$	66.67	67.63	72.89	71.21	53.65	62.86	73.58	71.21
	$S_5$	62.28	66.97	70.12	69	54.21	59.82	71.66	68.95

Table 6.2: Overall accuracies imagined speech, 3 features (vector-based approach).

		ORIGINAL				DWT				WPT			
		MLP	SVM	SSN <sub>L</sub>	SSN <sub>I</sub>	MLP	SVM	SSN <sub>L</sub>	SSN <sub>I</sub>	MLP	SVM	SSN <sub>L</sub>	SSN <sub>I</sub>
$\pm$ Nasal	$S_1$		62.91	67.65	68.13		63.02	66.16	66.55		63.02	65.9	66.82
	$S_2$		62.6	68.19	67.16		62.64	65.34	65.91		62.64	65.13	67.51
	$S_3$		62.63	74.2	65.36		62.68	68.5	66.76		62.68	68.02	67.39
	$S_4$		62.82	66.27	64.62		62.86	68.76	65.35		62.86	64.51	65.46
	$S_5$		61.16	64.7	62.94		61.19	68.04	68.29		61.19	63.77	67.03
$\pm$ Bilabial	$S_1$	58.62	64.32	69.72	69.67	51.81	62.85	65.06	68.39	51.62	62.85	65.06	68.17
	$S_2$	57.98	64.92	69.86	72.34	56.21	64.32	69	72.12	54.39	62.12	66.96	70.53
	$S_3$	59.71	62.9	75.67	74.03	55.7	62.95	65.41	66.37	54.39	62.95	65.65	66.61
	$S_4$	58.37	63.81	77.79	69.72	59.37	62.86	68.11	68.12	58.95	62.86	65.97	68.58
	$S_5$	56.97	59.78	66.9	67.69	57.52	59.82	66.01	65.82	55.75	59.82	62.51	66.51

On the other hand, Tables 6.2 and 6.3 show the overall accuracies obtained with imagined speech samples using 3 and 21 features, respectively, for the input vectors.

Due to the processing step with wavelets, imagined speech results were split into two tables, in which the "Original" column referred to when features were extracted from the EEG signal without wavelet processing. While the other columns represent the scores obtained with vectors composed of PCA components from DWT or WPT feature vectors.

Besides, similar to overt speech experiments, the best accuracies obtained were with  $\pm$ Bilabial data sets and SSN<sub>L</sub> classifier. However, these best results, based on the data set used, varied for the other classifiers, particularly for the SSN<sub>I</sub> in which  $\pm$ Nasal data set provided better accuracies in some cases.

Also, the best scores obtained with 3 and 21 features came from original signals (no wavelets usage). These results might mean that the mother wavelet was not adequate for these data and further experiments with other wavelet mothers are required.

Table 6.3: Overall accuracies imagined speech, 21 features (vector-based approach).

		ORIGINAL				DWT				WPT			
		MLP	SVM	SSN <sub>L</sub>	SSN <sub>I</sub>	MLP	SVM	SSN <sub>L</sub>	SSN <sub>I</sub>	MLP	SVM	SSN <sub>L</sub>	SSN <sub>I</sub>
$\pm$ Nasal	$S_1$		62.91	66.73	65.33		63.02	69.36	67.52		63.02	69.49	68.39
	$S_2$		62.6	64.13	64.81		62.64	65.13	67.26		62.64	65.02	67.55
	$S_3$		62.63	64.31	65.99		62.68	67.88	70.43		62.68	72.69	67.42
	$S_4$		62.82	61.34	64.48		62.86	72.65	66.69		62.86	66.11	66.81
	$S_5$		61.16	63.38	63.16		61.19	65.03	64.78		61.19	65.66	64.53
$\pm$ Bilabial	$S_1$		62.73	68.14	71.61		62.85	68.83	68.92		62.85	69.4	70.14
	$S_2$		64.28	68.37	70.79		64.32	68.89	69.15		64.32	71.42	72.17
	$S_3$		62.9	77.02	74.56		62.95	67.9	66.91		62.95	68.69	67.9
	$S_4$		62.82	75.02	72.49		62.86	70.77	67.49		62.86	71.46	66.11
	$S_5$		59.78	67.3	66.18		59.82	63.36	65.91		59.82	63.69	67.71

Furthermore, comparing the performances obtained across the experiments, the classifier that provided the best accuracies was the SSN (the neuron model varied per case), followed by the SVM and at last the MLP. While comparing the feature's set used, similar to overt speech samples, the results using 3 and 21 features were similar. However, these results were worse than those obtained with overt speech samples in this approach (Vector-based).

## 6.2 Spatio-Temporal

In this section are analyzed the results obtained with the Spatio-temporal classifiers used in this work: NeuCube and SSN (with LIF and Izhikevich neuron models). Table 6.4 summarizes these results for overt and imagined speech samples. Notice that, as in the Vector-based approach, the best results (highlighted) were provided by a  $\pm$ Bilabial data set using the SSN<sub>L</sub> classifier.

Besides, the best scores provided by each classifier were obtained from the  $\pm$ Bilabial data set, except for the NeuCube by using overt speech samples in which  $\pm$ Nasal data set provided slightly better results than  $\pm$ Bilabial. Furthermore, in some cases, the SSN<sub>I</sub> provided better results than the SSN<sub>L</sub>. For this reason, the results from both classifier models are detailed and compared in the next section. In this section, just NeuCube results are analyzed.

Although grid searches were performed for the NeuCube, their accuracies, compared with the SSNs results, were the worst in these experiments. These low performances occurred due to the few grid searches performed, and the few parameters explored. In the previous chapter was explained that two grid searches (corresponding to  $\pm$ Nasal and  $\pm$ Bilabial data sets) per mental activity (overt and imagined) were performed twice each.

By contrast, the optimization process in the SSN was performed for all the case individually (corresponding to each cell in Table 6.4)). However, a grid search or an optimization process (similar to that with the SSN) per case was not possible with the NeuCube due to time consumption. The NeuCube has several parameters and processes that would have implied many computations out of scope for time limitations in this work if an optimization process had been performed.

Table 6.4: Overall accuracies overt and imagined speech (spatio-temporal approach).

		OVERT SPEECH			IMAGINED SPEECH		
		NeuCube	SSN <sub>L</sub>	SSN <sub>I</sub>	NeuCube	SSN <sub>L</sub>	SSN <sub>I</sub>
$\pm$ Nasal	$S_1$	58.69	71.06	66.02	55.97	64.27	65.85
	$S_2$	58.89	64.16	62.18	57.83	64.35	64.74
	$S_3$	57.09	66.53	67.06	56.93	67.41	64.59
	$S_4$	60.32	71.79	66.87	58.07	65.93	62.68
	$S_5$	61.21	69.47	65.38	55.32	62.67	62.78
$\pm$ Bilabial	$S_1$	58.69	68.65	72.43	58.49	72.03	69.93
	$S_2$	58.12	71.14	74.13	57.22	72.66	71.78
	$S_3$	56.63	74.73	71.76	58.15	77.57	73.85
	$S_4$	58.99	70.92	69.47	58.64	75.09	68.08
	$S_5$	55.01	73.21	66.18	54.63	70.43	65.49

Table 6.5: NeuCube accuracies for overt speech samples using  $\pm$ Nasal data set.

Subject	$S_1$		$S_2$		$S_3$		$S_4$		$S_5$	
	Training	Test	Training	Test	Training	Test	Training	Test	Training	Test
1	57.11	65.03	60.77	52.70	61.93	59.38	56.89	62.79	58.99	60.92
2	60.26	54.60	60.13	62.16	59.28	47.66	60.18	65.12	60.53	58.62
3	58.16	52.76	62.05	48.65	60.96	42.97	59.96	54.65	60.75	62.07
4	60.26	60.12	61.41	58.11	61.45	54.69	58.86	59.30	58.33	59.77
5	61.58	56.44	58.42	54.05	60.24	56.25	59.74	55.81	58.77	57.47
6	56.05	48.47	61.19	59.46	62.89	47.66	61.71	59.30	60.31	54.02
7	63.42	61.35	58.21	58.11	65.78	41.41	61.49	55.81	64.25	62.07
8	61.84	53.99	59.28	56.76	65.54	42.19	58.64	56.98	61.62	68.97
9	59.21	49.69	62.05	59.46	58.07	50.78	64.99	58.14	62.72	62.07
10	62.37	60.12	62.26	50.00	64.58	52.34	56.89	63.95	58.99	64.37
11	57.63	58.90	62.47	54.05	60.72	53.13	63.24	54.65	64.25	59.77
12	58.68	61.96	62.47	58.11	61.45	58.59	61.27	65.12	61.62	64.37
13	54.47	52.76	60.98	62.16	63.86	52.34	62.36	55.81	59.43	57.47
14	60.00	65.64	60.98	68.92	61.93	56.25	59.30	58.14	60.31	64.37
15	61.32	63.19	59.49	62.16	64.58	59.38	62.36	65.12	59.21	60.92
16	57.37	58.28	59.70	63.51	58.80	49.22	61.49	65.12	58.33	60.92
17	60.53	56.44	62.26	62.16	63.86	54.69	61.71	55.81	61.18	64.37
18	60.26	61.35	57.78	58.11	61.69	58.59	62.58	56.98	62.28	63.22
19	58.16	53.37	57.14	48.65	60.48	51.56	59.96	65.12	64.04	57.47
20	60.53	63.80	56.50	52.70	61.20	55.47	60.18	65.12	59.21	70.11
Avg.	$59.46 \pm 2.25$	$57.91 \pm 5.05$	$60.28 \pm 1.88$	$57.5 \pm 5.36$	$61.96 \pm 2.18$	$52.23 \pm 5.6$	$60.69 \pm 2.03$	$59.94 \pm 4.19$	$60.76 \pm 1.95$	$61.67 \pm 3.88$

Also, Table 6.5 shows each classification accuracy, as well as their averages  $\pm$  the standard deviation, obtained with 20 different initializations for the NeuCube using overt speech samples and configured with the  $\pm$ Nasal data set (the best results obtained for this case, according to the overall accuracies shown in Table 6.4).

On the other hand, Tables 6.6 and 6.7 show the same information as in Table 6.5 but from the experiments that used  $\pm$ Bilabial configured imagined speech samples with the highlighted parameters for overt and imagined speech samples in Table 5.4, respectively.

Hence, the overall accuracies shown in Table 6.4 came from the Table 6.6 values because they were higher than those from Table 6.7. Furthermore, the most balanced cases of training and test accuracies with the NeuCube were selected from Tables 6.5 and 6.6 (those highlighted) to show their firing activities and weight's distributions.

Table 6.6: NeuCube accuracies for imagined speech samples using  $\pm$ Bilabial data set (1st parameter set).

Subject	$S_1$		$S_2$		$S_3$		$S_4$		$S_5$	
	Init.	Training	Test	Training	Test	Training	Test	Training	Test	Training
1	61.05	52.76	65.46	55.41	56.14	57.81	63.68	58.14	63.16	42.53
2	60.79	57.67	60.13	56.76	60.72	60.94	61.49	45.35	59.65	47.13
3	59.21	51.53	61.19	56.76	58.55	64.06	58.42	59.30	65.35	47.13
4	57.11	49.08	61.19	43.24	61.20	53.91	59.08	60.47	64.04	47.13
5	61.32	38.65	62.05	51.35	56.14	65.63	61.27	51.16	64.47	49.43
6	62.11	58.28	62.69	59.46	57.11	58.59	62.58	52.33	62.94	48.28
7	64.47	60.74	60.98	55.41	53.98	64.84	59.96	55.81	66.45	50.57
8	60.79	53.37	60.34	64.86	57.11	55.47	58.42	63.95	60.75	50.57
9	59.47	64.42	60.55	55.41	59.04	58.59	63.24	53.49	62.06	41.38
10	64.47	65.64	61.83	40.54	60.00	57.81	59.30	59.30	62.28	44.83
11	60.79	66.26	63.33	48.65	61.45	59.38	58.42	56.98	61.40	50.57
12	63.68	60.12	66.10	56.76	58.31	63.28	60.61	58.14	63.16	49.43
13	64.47	53.99	57.57	43.24	56.87	58.59	61.93	55.81	62.50	48.28
14	64.21	46.63	63.75	59.46	56.14	53.91	63.68	50.00	60.09	43.68
15	60.53	46.63	62.26	37.84	58.07	61.72	62.80	52.33	62.72	52.87
16	65.26	62.58	62.05	55.41	59.52	55.47	58.64	58.14	60.75	48.28
17	62.11	52.76	61.62	58.11	62.17	46.09	62.14	54.65	62.06	49.43
18	61.32	49.69	62.90	67.57	58.07	61.72	61.49	60.47	62.06	37.93
19	64.21	55.83	60.98	50.00	56.14	51.56	61.49	60.47	63.38	36.78
20	58.16	57.67	60.55	35.14	56.63	53.13	57.77	62.79	63.82	45.98
Avg.	<b>61.78<math>\pm</math>2.32</b>	<b>55.21<math>\pm</math>7.13</b>	<b>61.88<math>\pm</math>1.89</b>	<b>52.57<math>\pm</math>8.73</b>	<b>58.17<math>\pm</math>2.16</b>	<b>58.12<math>\pm</math>4.93</b>	<b>60.82<math>\pm</math>1.93</b>	<b>56.45<math>\pm</math>4.64</b>	<b>62.65<math>\pm</math>1.7</b>	<b>46.61<math>\pm</math>4.28</b>

Table 6.7: NeuCube accuracies for imagined speech samples using  $\pm$ Bilabial data set (2nd parameter set).

Subject	$S_1$		$S_2$		$S_3$		$S_4$		$S_5$	
	Init.	Training	Test	Training	Test	Training	Test	Training	Test	Training
1	62.11	52.15	62.26	55.41	55.66	60.16	62.58	61.63	61.62	42.53
2	65.79	50.92	58.00	60.81	53.49	49.22	57.33	52.33	64.04	42.53
3	62.63	52.15	60.13	54.05	60.72	60.94	63.68	51.16	59.87	44.83
4	62.89	56.44	61.83	59.46	58.31	54.69	59.08	58.14	62.28	45.98
5	63.68	52.76	61.83	52.70	59.04	59.38	62.36	58.14	59.21	43.68
6	63.42	54.60	62.47	44.59	57.11	58.59	61.27	61.63	59.87	44.83
7	61.05	60.12	62.69	51.35	57.59	56.25	59.74	51.16	67.32	42.53
8	60.53	33.74	60.13	51.35	60.48	56.25	60.39	52.33	64.69	47.13
9	70.00	39.26	58.64	55.41	60.00	56.25	61.27	60.47	62.06	50.57
10	60.26	61.35	56.08	45.95	60.24	60.94	62.58	53.49	62.72	39.08
11	60.79	37.42	61.83	52.70	56.63	54.69	63.02	54.65	65.13	48.28
12	62.37	55.21	60.55	36.49	57.11	64.06	63.68	43.02	58.55	36.78
13	62.63	64.42	64.39	52.70	56.39	58.59	63.89	56.98	63.38	35.63
14	61.32	44.17	62.05	44.59	59.52	60.16	58.86	48.84	62.06	43.68
15	59.74	57.67	60.77	52.70	58.55	59.38	62.58	58.14	59.43	50.57
16	56.84	44.79	61.41	64.86	59.28	54.69	60.39	61.63	59.21	42.53
17	63.95	63.80	65.88	43.24	60.00	57.81	61.71	58.14	62.06	42.53
18	60.79	35.58	58.85	52.70	57.11	57.81	60.61	59.30	63.16	43.68
19	62.11	53.37	55.86	54.05	59.76	58.59	59.30	53.49	65.35	45.98
20	63.68	58.28	60.55	47.30	53.25	63.28	60.39	55.81	59.65	55.17
Avg.	<b>62.3<math>\pm</math>2.62</b>	<b>51.41<math>\pm</math>9.29</b>	<b>60.81<math>\pm</math>2.48</b>	<b>51.62<math>\pm</math>6.56</b>	<b>58.01<math>\pm</math>2.18</b>	<b>58.09<math>\pm</math>3.37</b>	<b>61.24<math>\pm</math>1.82</b>	<b>55.52<math>\pm</math>4.86</b>	<b>62.08<math>\pm</math>2.44</b>	<b>44.43<math>\pm</math>4.6</b>

Figures 6.1 and 6.2 show, the initial and final (post-training) firing activities from each NeuCube connection by feeding them with overt and imagined speech samples, respectively.

In both cases, the firing activities at the top of the plots correspond to the connections between input neurons and other reservoir neurons, while those from the bottom represents the output connections. For this reason, such firing activities were the most saturated, even with the initial weights (top plots from Figures 6.1 and 6.2).

Besides, in both figures seems that those connections with initial high firing activity became more saturated after training the NeuCube. On the other hand, those connections with few or no activity in the initial step remained with no firing activity after the training step, which means that the information did not propagate through those connections.

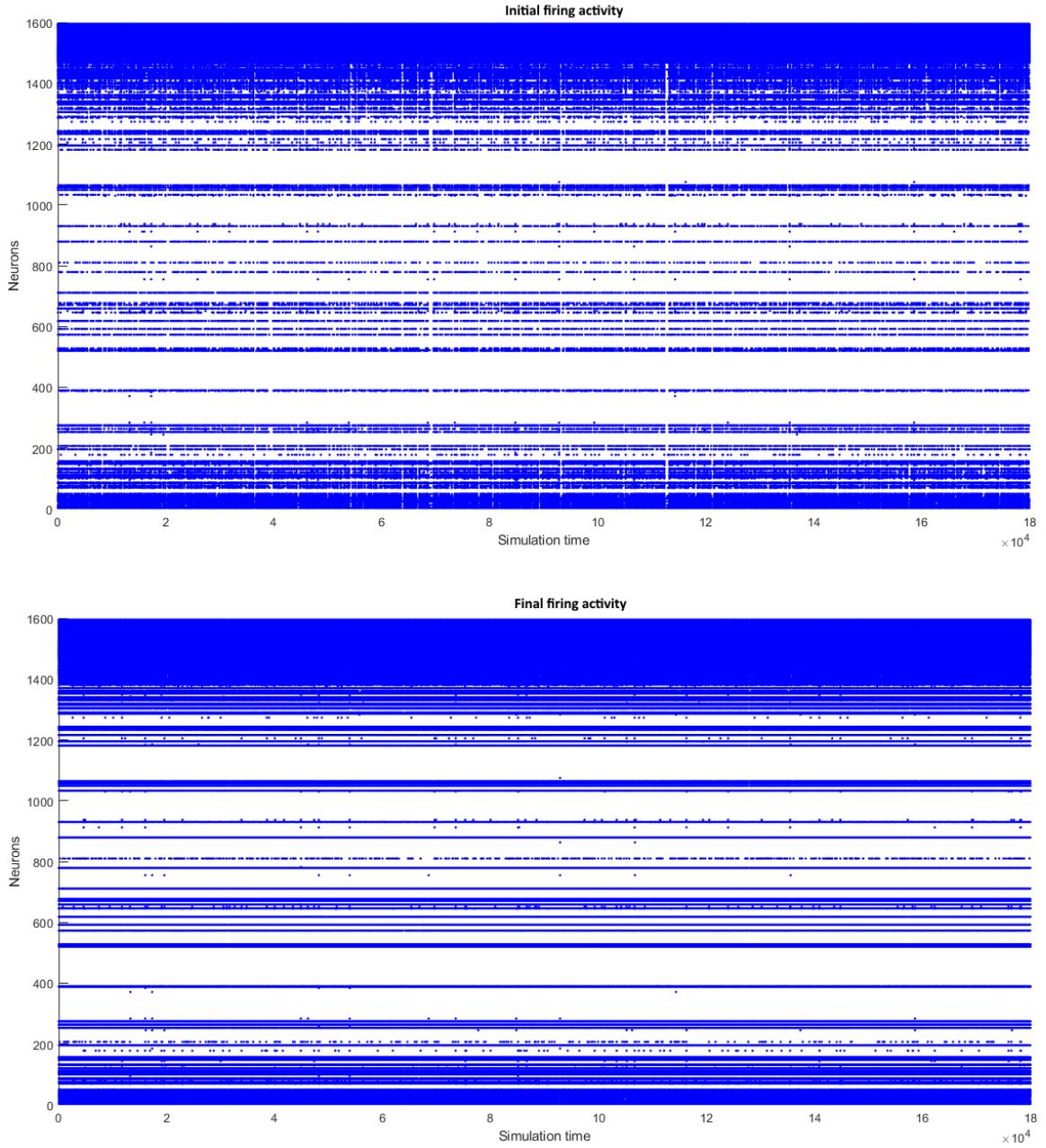


Figure 6.1: Initial and final firing activity with the NeuCube using overt speech samples.

Moreover, Figures 6.3 and 6.4 show, respectively, for overt and imagined speech samples the initial and final weight's distributions from the same highlighted cases in Tables 6.5 and 6.6. Thus, at the top of the distributions are shown the amount of positive ( $p$ ), negative ( $n$ ), and zero ( $z$ ) weight's values.

Notice that the initial distributions in both Figures were similar due to the value selected for  $r_+$ , which set 30% of the connections as excitatory (positives) and 70% as inhibitory (negatives). Hence, in both Figures, the final distribution shows that most of the weight's values are close to zero, except for some weights that presented values of -2.

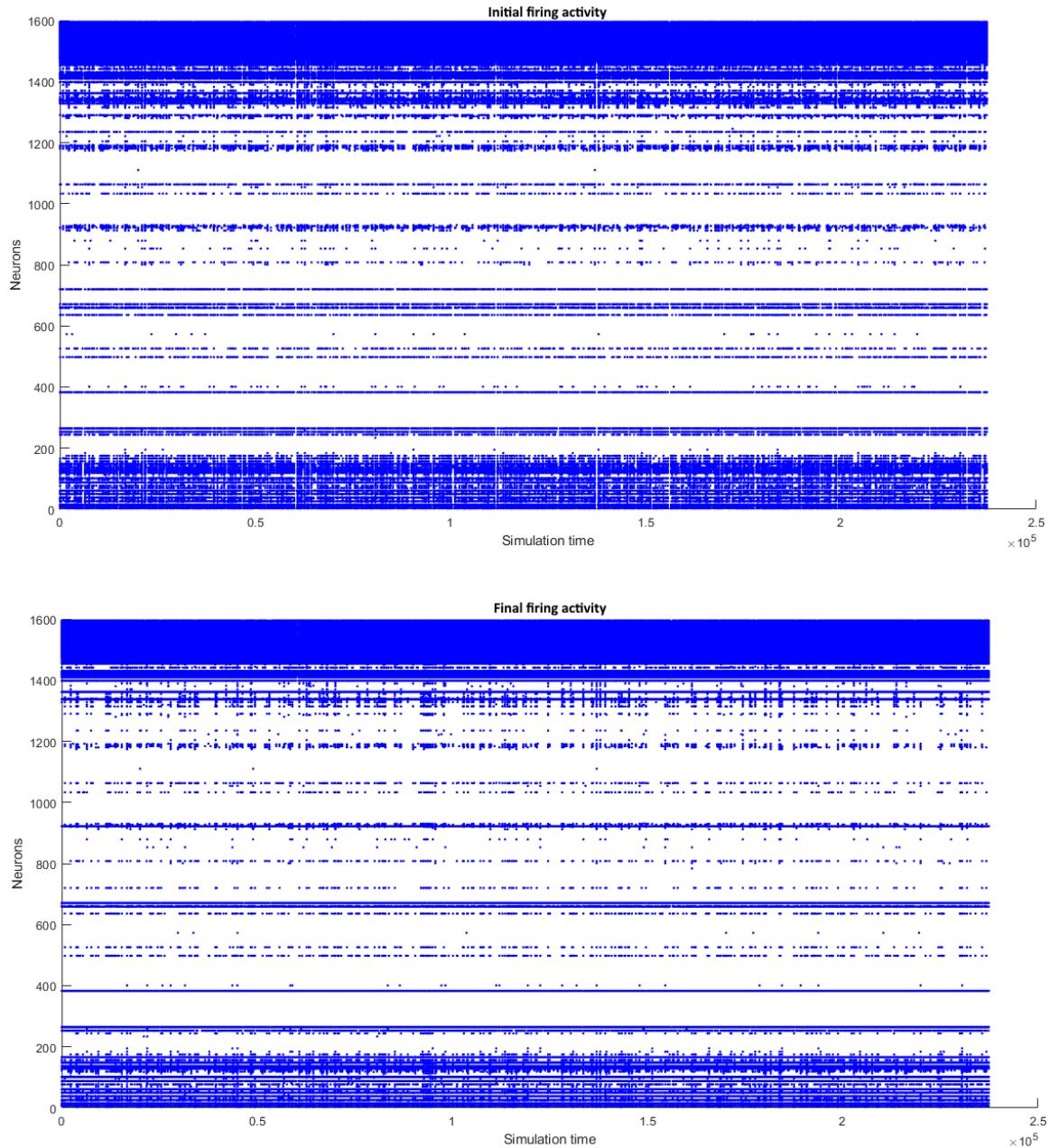


Figure 6.2: Initial and final firing activity with the NeuCube using imagined speech samples.

Furthermore, in Figures 6.5 and 6.6 are shown the initial (pre-training) and final (post-training) weight's connections for the cases exposed here of overt and imagined speech samples, respectively.

In both Figures, the green points represent the input neurons, while the blue points are the reservoir neurons. Indeed, all neuron's locations are based on the Talareich template, which has a total of 1471 neurons.

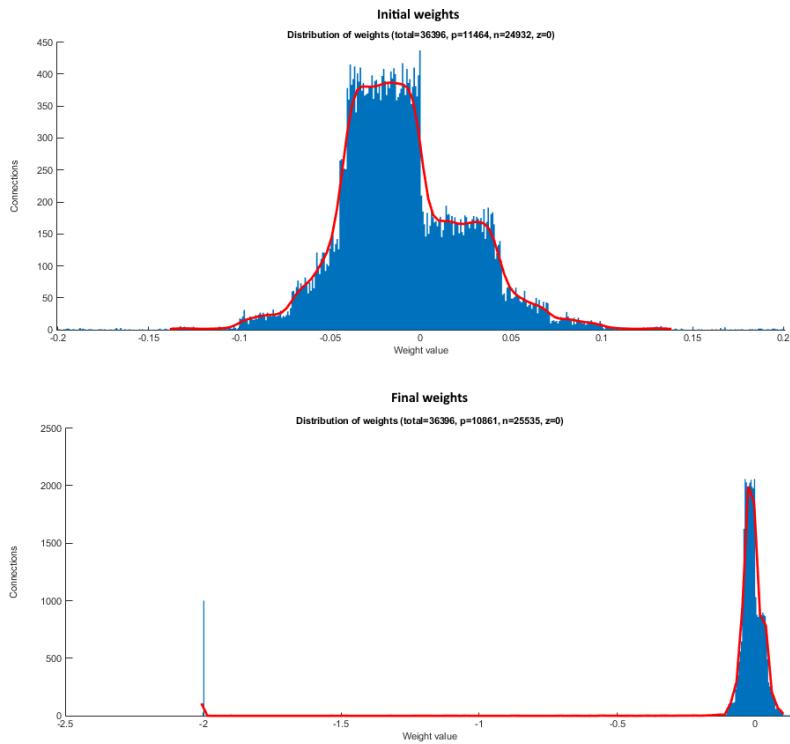


Figure 6.3: Initial and final weights with the NeuCube using overt speech samples.

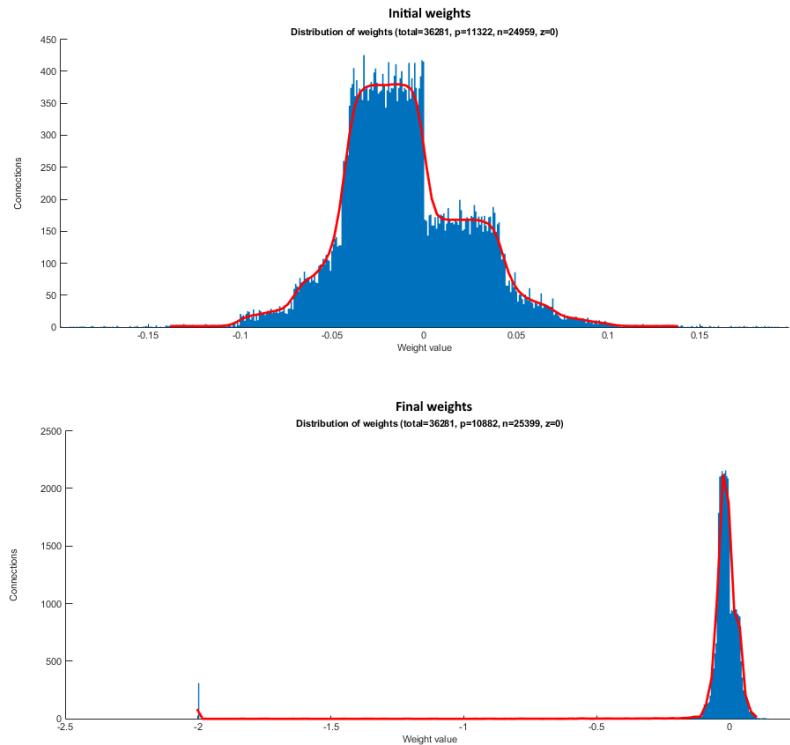
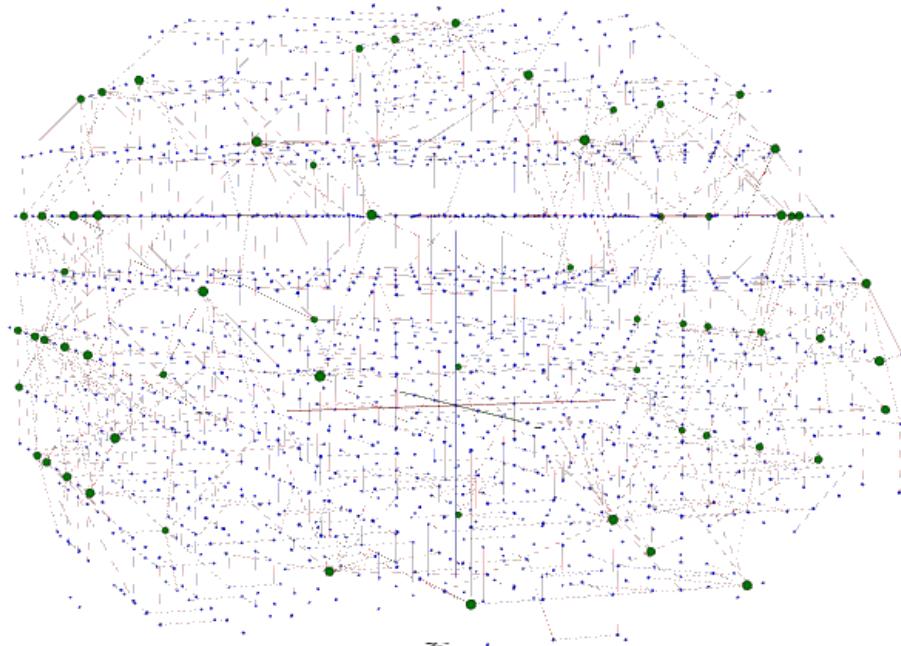


Figure 6.4: Initial and final weights with the NeuCube using imagined speech samples.

Initial weight connections



Final weight connections

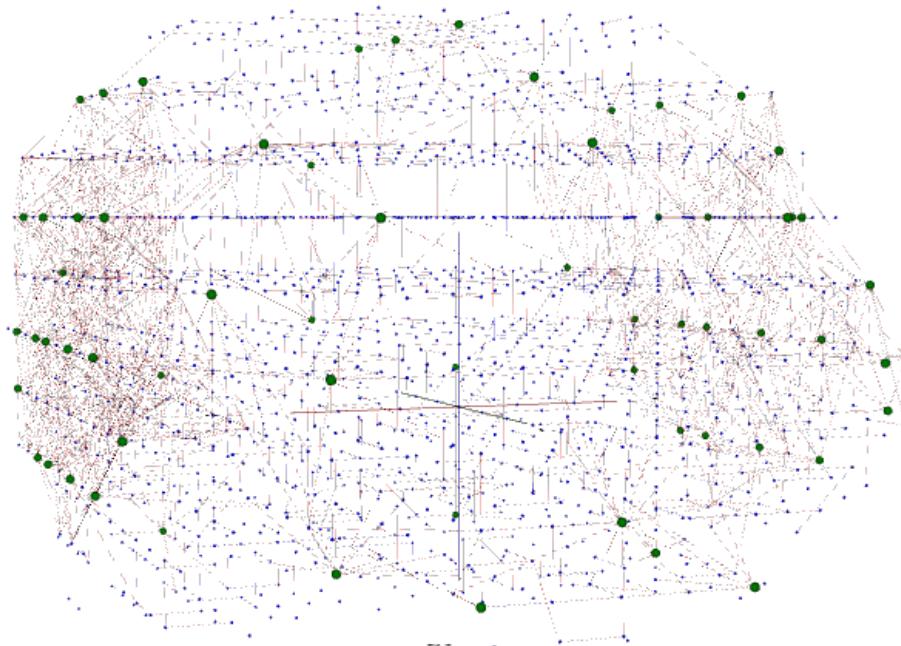
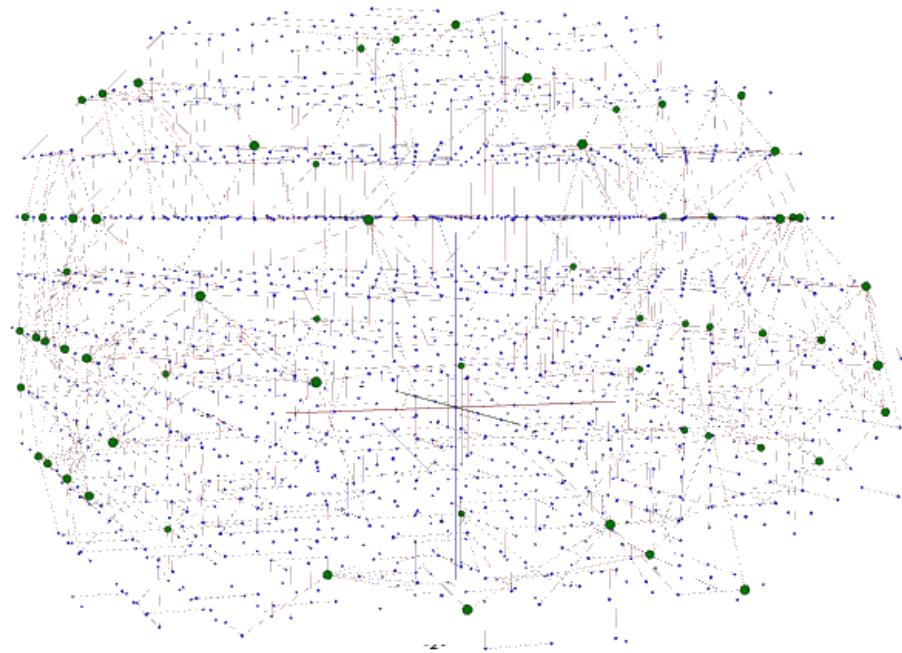


Figure 6.5: Initial and final NeuCube connections for overt speech samples.

Initial weight connections



Final weight connections

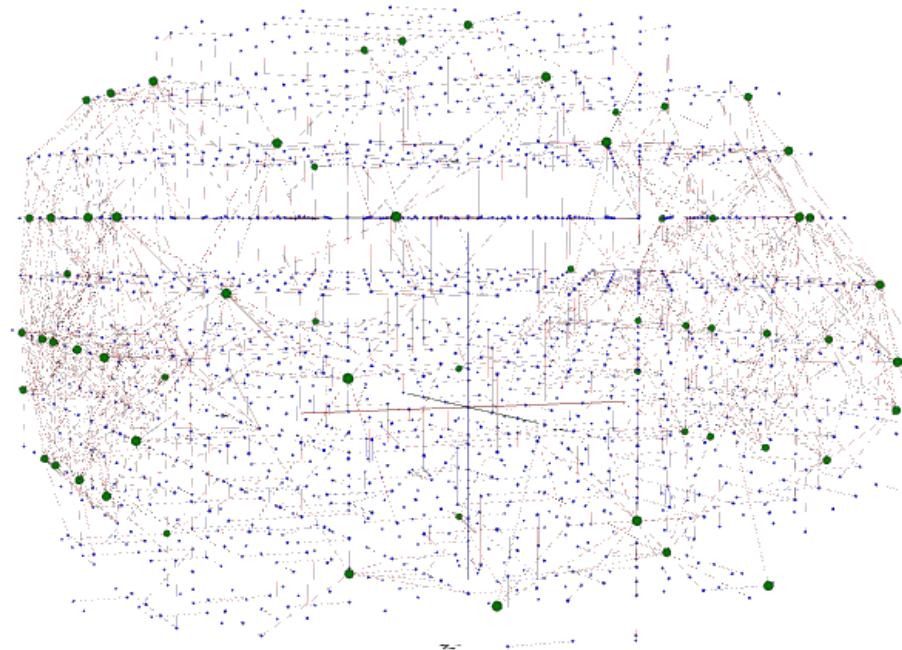


Figure 6.6: Initial and final NeuCube connections for imagined speech samples.

Hence, the blue and red lines represent the excitatory and inhibitory connections, respectively. Notice that there are more inhibitory than excitatory connections in both Figures, particularly on the initial reservoir, which match with the weight's distributions shown in Figures 6.3 and 6.4.

Also, in both Figures, more connections are represented in the final reservoir plots than in the initial. These behaviors happened because just the connections with weights below -0.7 and above 0.7 are shown, and, according to Figures 6.3 and 6.4, the initial weight's distribution was sparser than the final, which caused more weights discrimination when the interval was selected.

To sum up, the firing activities from Figures 6.1 and 6.2 shown that the information did not pass through the reservoir, just in the input connections and some few others. This argument is reinforced with the final connections shown in Figures 6.5 and 6.6. Due to that, the NeuCube has failed in many recognitions for training and test data. Besides, the grid search was performed with few cases, with which was not possible to fully explored the NeuCube capacities in this work.

### 6.3 Mixed Approach Analysis

As the SSN classifier provided the best scores, in this section is made a more in-depth analysis of these results. Hence, Figures 6.7 and 6.8 show, respectively, the firing activity of the same overt and imagined speech cases represented Figure 4.15 (errors obtained per generation). These cases also correspond to the 21 feature based case highlighted in Table 6.3 from the  $S_3$  test data for the case of overt speech, and to the highlighted case from  $S_3$  test data in Table 6.4 for the case of imagined speech.

Furthermore, the firing activities showed above the horizontal black line in both Figures 6.7 and 6.8 came from samples of class +Bilabial, while those below the line came from the other class -Bilabial. Besides, the misclassified samples are represented with the red colored firing activities. Thus, it can be seen in Figure 6.7 that +Bilabial samples were more commonly misclassified than -Bilabial samples. On the other hand, in Figure 6.8 were presented similar amounts of errors in both class sets.

Due to that, the error percentage was calculated per class in each case, which is closely related to the classification metrics and represents the amount of misclassified samples divided by the total number of samples in that class. Tables 6.8 and 6.9 show these error percentages, as well as the averages, obtained with the best configurations (according to the highlighted overall accuracies in Tables 6.1, 6.2, 6.3, and 6.4) using overt and imagined speech samples, respectively.

It can be seen in both Tables 6.8 and 6.9 that, on average, -Bilabial samples were recognized better than +Bilabial samples in all the experiments. From the particular cases shown in Figures 6.7 and 6.8, the error percentages were 2.65% and 73.89% for overt speech example; while the error percentages were 25.59% and 34.48% for imagined speech example.

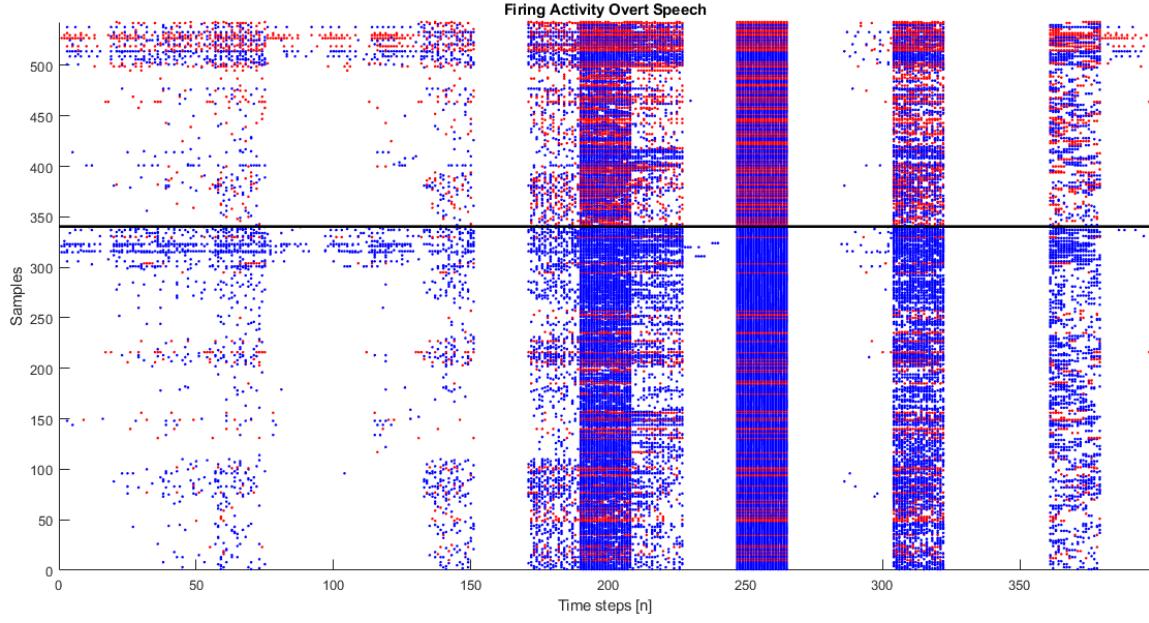


Figure 6.7: Firing activity per sample class and misclassified samples highlighted after SSN testing for overt speech example.

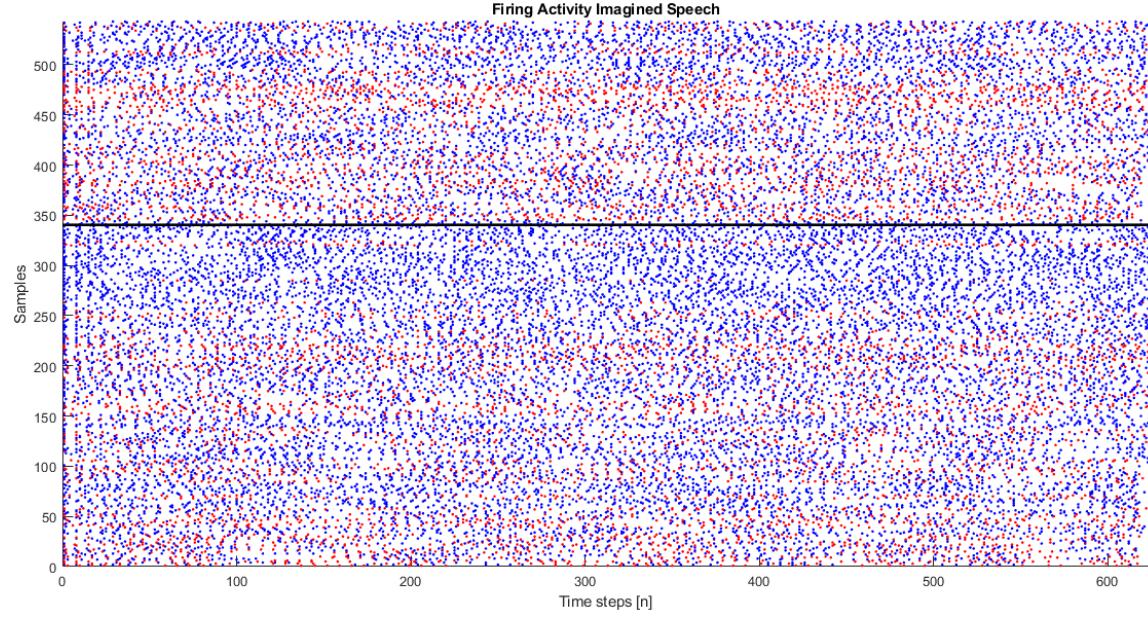


Figure 6.8: Firing activity per sample class and misclassified samples highlighted after SSN testing for imagined speech example.

Besides, in the case of overt speech, the AFR associated to the class –Bilabial and +Bilabial were 0.1514 and 0.18.66, respectively; while for imagined speech, the AFR for class –Bilabial and +Bilabial were 0.0694 and 0.0725, respectively. Also, Figures 6.9 and 6.10 show in different plots, for overt and imagined speech classification respectively, training and test accuracies (percentages) across the experiments by using  $SSN_L$  and  $SSN_I$  classifiers.

Table 6.8: Error percentages from the best results using SSN for overt speech.

Subjects	Vector-based				Spatio-temporal	
	3 features		21 features			
	-Bilabial	+Bilabial	-Bilabial	+Bilabial	-Bilabial	+Bilabial
$S_1$	12.94	46.31	2.36	69.95	10.88	65.03
$S_2$	14.12	44.34	2.36	69.46	9.71	56.65
$S_3$	7.65	55.67	2.65	73.89	8.82	57.64
$S_4$	6.76	57.64	2.36	70.44	8.53	57.14
$S_5$	9.12	62.56	2.06	69.95	11.47	49.75
Avg.	$10.12 \pm 3.25$	$53.30 \pm 7.74$	$2.36 \pm 0.21$	$70.74 \pm 1.8$	$9.88 \pm 1.28$	$57.24 \pm 5.41$

Table 6.9: Error percentages from the best results using SSN for imagined speech.

Subjects	Vector-based				Spatio-temporal	
	3 features		21 features			
	-Bilabial	+Bilabial	-Bilabial	+Bilabial	-Bilabial	+Bilabial
$S_1$	16.47	42.37	17.40	50.74	16.47	48.28
$S_2$	15.00	43.84	18.29	49.75	20.88	41.38
$S_3$	15.59	41.38	20.06	44.83	25.59	34.48
$S_4$	15.88	41.38	16.81	50.25	21.18	38.42
$S_5$	13.82	47.78	10.91	59.61	23.24	35.47
Avg.	$15.35 \pm 1.01$	$43.35 \pm 2.68$	$16.69 \pm 3.46$	$51.03 \pm 5.35$	$21.47 \pm 3.37$	$39.61 \pm 5.55$

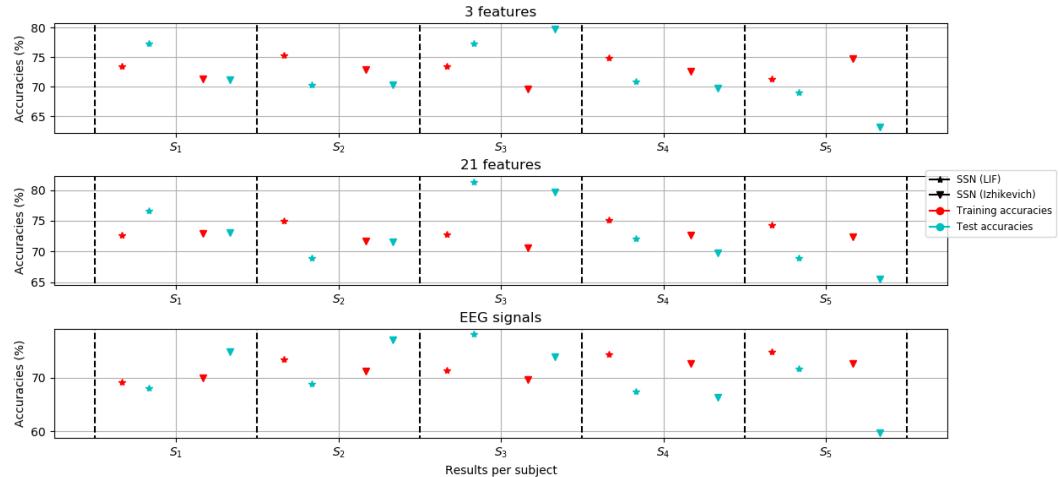


Figure 6.9: Accuracies of the best configurations per case for overt speech samples.

In general,  $SSN_L$  accuracies outperformed those obtained with  $SSN_I$  for training and test data. Hence, the training accuracies were higher than those from test data in almost all cases. However, it happened the opposite in both Figures when  $S_3$  data (and other subject's data in some particular cases) were held out.

These results mean that the classifiers could generalize better both mental activities with the rest data than with  $S_3$  data, even though the training accuracies were lower than when other subject's data different from  $S_3$  were held out.

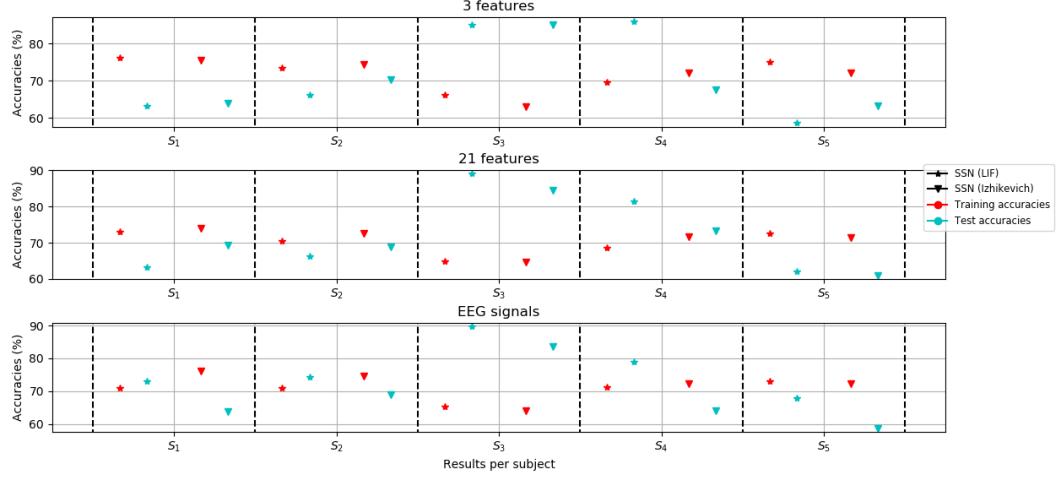


Figure 6.10: Accuracies of the best configurations per case for imagined speech samples.

Table 6.10: Best scores obtained with  $SSN_L$  for overt speech samples.

	Subject	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	Avg.
Training	<b>Accuracy</b>	73.42	75.27	73.49	74.84	71.27	<b>73.66±1.56</b>
	<b>Recall</b>	76.06	77.02	73.42	73.48	70.94	<b>74.18±2.41</b>
	<b>Precision</b>	83.47	85.52	89.92	93.33	93.13	<b>89.07±4.45</b>
	<b>F1</b>	79.60	81.05	80.84	82.23	80.53	<b>80.85±0.95</b>
Test	<b>Accuracy</b>	77.30	70.27	77.34	70.93	68.97	<b>72.96±4.04</b>
	<b>Recall</b>	75.57	73.33	73.87	70.83	70.37	<b>72.8±2.17</b>
	<b>Precision</b>	95.19	88.00	100.00	92.73	77.55	<b>90.69±8.53</b>
	<b>F1</b>	84.26	80.00	84.97	80.31	73.79	<b>80.67±4.45</b>

Table 6.11: Best scores obtained with  $SSN_L$  for imagined speech samples.

	Subject	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	Avg.
Training	<b>Accuracy</b>	71.05	71.00	65.30	71.12	73.03	<b>70.3±2.92</b>
	<b>Recall</b>	76.25	76.74	73.95	77.22	80.00	<b>76.83±2.17</b>
	<b>Precision</b>	77.54	76.21	68.22	76.14	76.98	<b>75.02±3.84</b>
	<b>F1</b>	76.89	76.47	70.97	76.68	78.46	<b>75.89±2.86</b>
Test	<b>Accuracy</b>	73.01	74.32	89.84	79.07	67.82	<b>76.81±8.31</b>
	<b>Recall</b>	71.13	73.85	90.59	78.46	69.81	<b>76.77±8.41</b>
	<b>Precision</b>	97.12	96.00	93.90	92.73	75.51	<b>91.05±8.86</b>
	<b>F1</b>	82.11	83.48	92.22	85.00	72.55	<b>83.07±7.06</b>

Moreover, the plots from Figure 6.9 corresponding to Vector-based experiments presented similar results. For this reason and due to less required computations, the scores from 3 feature vector experiments with  $SSN_L$  were considered in this work as the best obtained for overt speech samples. On the other hand, the scores from Spatio-temporal experiments with  $SSN_L$  were regarded as the best obtained for imagined speech samples.

Tables 6.10 and 6.11 show in detail these best scores obtained from overt and imagined speech experiments, respectively. These scores consisted of the accuracy, recall, precision, and F1 per subject's data held out from the rest.

Besides, the averages±standard deviations are shown across the experiments per score, which show, for both mental activities, that the variations in all training scores remained low, while test scores variations were high (mainly because of  $S_3$  data results). These low variations in the training scores show that, for both mental activities, the  $SSN_L$  learning capacity is similar when different training samples are used.

Finally, Figures 6.11 and 6.12 show the confusion matrices from which the best scores of imagined and overt speech, respectively, were extracted. In the case of overt speech (Figure 6.11), almost all the samples (correctly or incorrectly) were classified as –Bilabial (the range values are from 0 to 1). Although the least unbalanced binary configuration was used, the amount of –Bilabial samples (340 against 203) aided the scores to be higher than 70%.

On the other hand, in the case of imagined speech training samples (first column of Figure ??), the recognition rate of both classes was balanced and slightly high on average (76% for –Bilabial and 60% for +Bilabial). While, the same situation as for overt speech happened with test samples, except when  $S_3$  data was used, which provided high recognition rates for both classes (94% for –Bilabial and 83% for +Bilabial).

To sum up, the classification rates for both mental activities were above 70%, although the unbalance between class samples aided to achieve these results. Besides, imagined speech classification rates outperformed those from overt speech, possibly due to the amount of temporal information used in the SSN (625 sample points against 57 features). These results agreed with those from [14], in which the best scores for imagined speech were obtained with EEG raw data. Hence, these results show the importance of designing new acquisition experiments, as well as the necessity of collecting new balanced samples. These aspects are summarized in the next chapter.

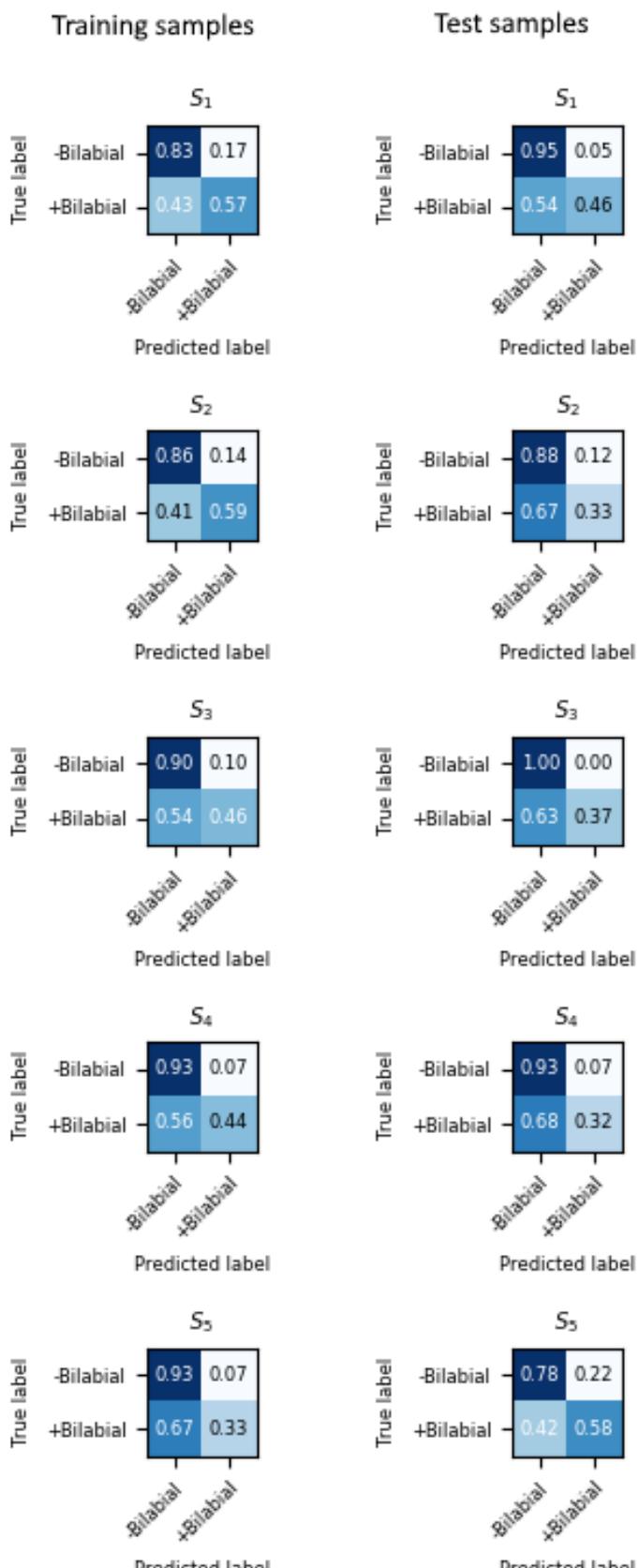


Figure 6.11: Confusion matrices from best overt speech scores.

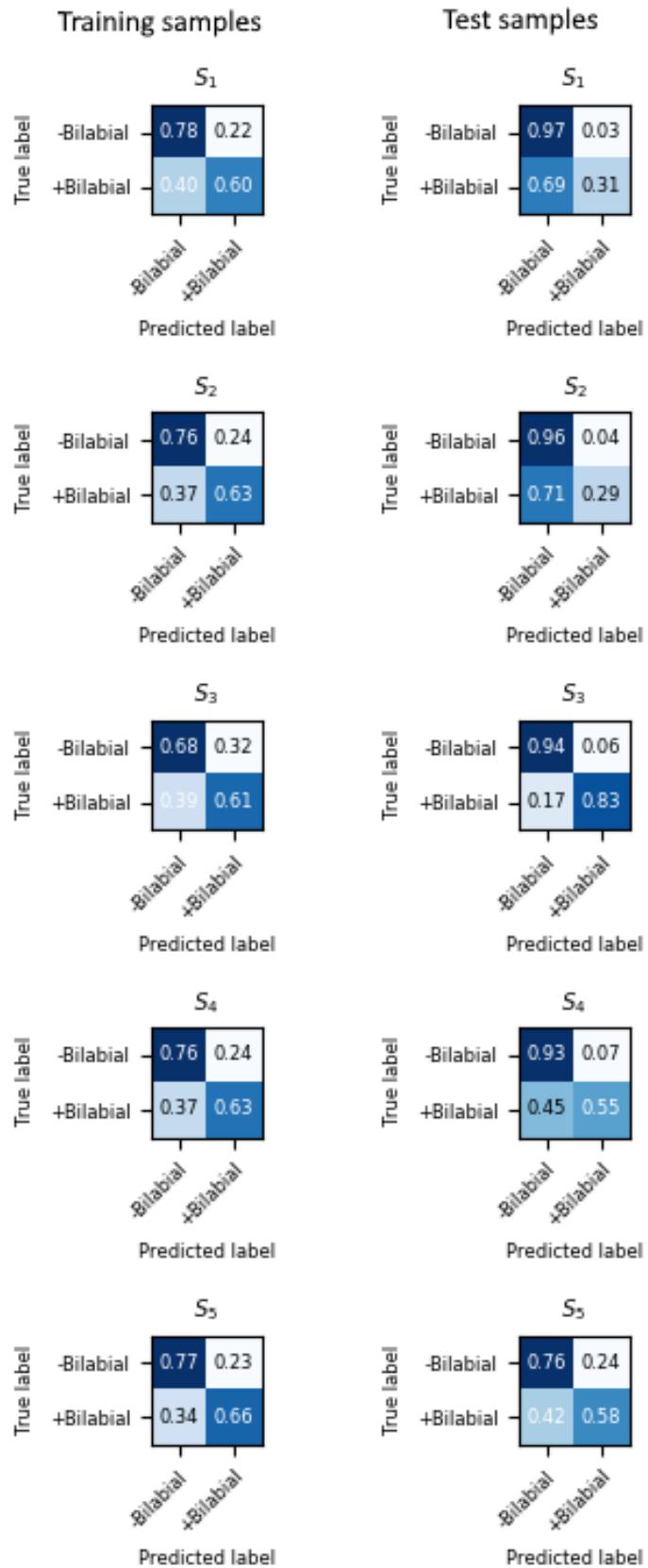


Figure 6.12: Confusion matrices from best imagined speech scores.

# CONCLUSIONS AND FUTURE WORK

---

In this chapter are described the conclusions associated with the particular objectives of this work. Due to that, each subsection is related to a particular objective. Hence, the most outstanding contributions from each step of the methodology followed (Figure 4.1) are mentioned. Finally, the detected future works from the methodology steps are mentioned at the end of this chapter.

## 7.1 Conclusions

### 7.1.1 Data Selection and Processing

In chapter 4 was mentioned that all database's signals were revised and, consequently, some of them were discarded due to notorious periodicities presented on the signals. With this step was remarked the importance of revising the samples before performing experiments, despite the database's usage in previous works. Besides, more samples are required to obtain more significant results (here were used a subset of 543 from the initial amount of 1870), as well as a balance across the number of class samples is necessary.

Furthermore, the processing steps carried in this work were specific for each classification approach and mental activity samples. In the Vector-based approach, the processing was different for each mental activity. For overt speech samples, the signals were trimmed according to speech signals energy. Whereas, for imagined speech samples, two wavelet approaches (DWT and WPT) were used to perform a spectro-temporal analysis over the signals (due to the absence of external references). Besides, both wavelet approaches coincided with those EEG sub-bands described in the theoretical framework.

On the other hand, the processing step in the NeuCube (Spatio-temporal approach) was the same for both mental activities and consisted of encoding the EEG signals. This step was carried with optimization processes (grid search for SF and DEA for BSA) through the MSE measured with the signal's reconstruction. While the SSN did not require necessarily any processing step (just the preprocessed signals, and in the case of feature vectors, the same processing as in the Vector-based approach).

### 7.1.2 Feature Extraction

The features extracted from the signals were particular for each classification approach. In the case of the Vector-based approach, statistical features were computed from each signal's overlapped segment, and PCA was performed when wavelets were applied in the processing step to reduce the vector's dimensionality. These statistical features were input to the classifiers (MLP and SVM in this work) that cannot work with raw temporal data.

On the other hand, the feature extraction step in the NeuCube can be considered when the optimal encoding of the signal was obtained. While the SSN did not require any particular feature extraction since it accepted the input values (either features or signal's amplitude) each time step.

### 7.1.3 Experimental Framework

Next, the validation methodology followed in this work considered subject-independent experiments by holding one subject's data out from the rest per case. Hence, training and test scores were computed with data unseen previously by the trained classifiers, which provided unbiased results. Furthermore, several experiments with different initializations were performed for classifiers sensitive to the initial parameter's values (MLP and NeuCube in this work). Then, the average scores were computed to provide more reliable results.

### 7.1.4 Classification Results Analysis

In chapter 6 were shown all the overall accuracies obtained with each classifier, binary data set used, and subject's data held out from the rest for testing. Moreover, this information was presented by mental activity, classification approach, and features used (in the case of the Vector-based approach). Then, the best overall accuracies per case were selected to make a more in-depth analysis.

These best results proceeded from the  $\pm$ Bilabial data set and by the  $SSN_L$  classifier. Hence, analysis over the  $SSN_L$  results was carried, from which the best per mental activity were selected. For overt speech samples, the best scores were obtained with the Vector-based approach using 3 features, while for imagined speech were with the Spatio-temporal approach.

Then, the average was computed across training and test accuracies. For overt speech samples, the training and test accuracies were  $73.66 \pm 1.56\%$  and  $72.96 \pm 4.04\%$ , respectively. While for imagined speech samples, the average accuracies from training and test data were  $70.3 \pm 2.92\%$  and  $76.81 \pm 8.31$ , respectively.

These results were lower than those reported in the previous works that used the database Kara One (above 90% of accuracies, on average). However, they used the samples rejected in this work and were considered not examples of any mental activity. Hence, their periodicities aided to get similar features among the samples from the same class and, consequently, provided high scores.

Besides, the reported results in those works came from the most unbalanced binary classes:  $\pm VC$  and  $\pm uw/$ , which, as seen with the best results in this work, also aided to improve the accuracies. Additionally, in the case of the classifier that provided the best scores in those works (the DBN), it was not mentioned if several experiments with different initializations were performed, nor cross-validation for training scores as in this work.

On the other hand, comparing with other EEG-based works in the literature review, in this work were used classifiers that consider the temporal and spatial aspects of the data. In the Vector-based approach, the temporal information was extracted from each channel classifier (spatial aspect considered with them) and considered jointly with the decision rule, while in other related works, each channel data was averaged or concatenated, which masked each channel's contributions.

Hence, although the Pearson correlations between acoustic and imagined feature vectors were computed for the channel's selection in the previous works that used the Kara One database, a test to obtain the significance of these correlations were not carry (such as the Kolmogorov-Smirnov test). While, in this work, a channel's selection methodology was carried based on the best resulting scores. Besides, in the Spatio-temporal approach, were used classifiers, based on spiking neurons, that had not been used before for overt and imagined speech samples.

To sum up, a methodology was carried with Vector-based approach classifiers to deal with spatial and temporal information. Hence, Spatio-temporal classifiers based on spiking neurons have been used by the first time with overt and imagined speech, from which the  $SSN_L$  scores outperformed others with training and test accuracies above the chance in both mental activities. These results could represent the feasibility of overt and imagined speech classification. However, a more in-depth analysis showed that the unbalance between class samples aided to obtain high accuracy. For this reason, it is recommended to develop balanced databases in future works.

## 7.2 Contributions

The contributions from this work are listed as follows:

- Nosy samples rejected and indicated from the database Kara One for future works.
- Overt speech classification experiments (these samples had not been used in other previous works with this database).
- Classification experiments with Vector-based and Spatio-temporal approaches.
- Wavelet decompositions with two approaches (DWT and WPT). Hence, MODWPT was used to avoid downsampling (data reduction).
- Score's comparisons between the 21 features described in the literature review, and a subset of these features.
- Wavelet feature's dimension reduced with PCA.

- Sub-bands selection by correlation rankings (not used in all the classification experiments due to low scores).
- A voting scheme for Vector-based classifiers.
- Channel's selection, per mental activity and classifier, based on channel's scores.
- Optimized EEG-signal's encodings with two different encoders (SF and BSA).
- Subject-independent classification with the NeuCube for the first time.
- Two different neuron models used with the SSN: LIF and Izhikevich.
- Two different input types for the SSN: feature vectors and EEG signals.
- A subject-independent validation methodology for all the classifiers, which avoided biased results by computing the scores from data not used in training.
- Average scores computed from several experiments with different initializations for classifiers sensitives to initial parameters.
- Parameters used in each classifier were shown to replicate the experiments.
- Classification results analysis through the accuracy, recall, precision, and F1 scores.
- An article (not submitted yet) written during the research stay at the KEDRI lab from the AUT.

### 7.3 Future Recommendations

Due to the problems founded with the Kara One database, it is recommendable to create a new robust and balanced database, from which all samples were revised beforehand. Hence, more controlled data acquisition experiments are required, like those from [38], in which the exact timing intervals of both mental activities were recognized due to the visual stimulations. Thus, the signals from both mental activities were shorter than those from the Kara One database. Also, wavelet decompositions would be applied for both mental activities independently of their lengths, if EEG signals were recorded with a lower Fs.

On the other hand, a more in-depth study in the preprocessing techniques for EEG signals is required to implement them in future works since there were not any contributions to this step in this work. ICA [39] is a preprocessing method usually used. However, it is recommended to carry a well-based analysis of the resulting sources since overt and imagined speech phenomena currently are not entirely understood.

Moreover, wavelet decompositions were used in the processing step. However, the scores obtained with these techniques were low compared with no processed (imagined speech) signals. These results may reflect the inadequate mother wavelet selection because, as stated in the theoretical framework, a properly wavelet selection could provide optimum resolution of specific neuroelectric events. Hence, the matching pursuit method or matched Meyer wavelets (mentioned in [40]) are recommended for these data to create proper wavelets.

Then, in this work were explored just some statistical features and a subset from them. It is recommended to use other types, such as spectral features, in future works. Thus, a feature selection step could be useful to reduce the number of classification experiments. A bag of features can be used, similarly as in [14], as well as other methods to obtain the feature sets that maximize the distance between samples from different classes while minimizing the distance of samples from the same class.

Furthermore, other segment lengths and overlapping percentages could be tested, as in [8]. Due to that, other classifiers that consider temporal relationships could be used, such as the HMMs, which each could process a particular input channel data. Thus, each HMM could have transition states with the nearby HMMs, with which temporal and spatial aspects of the EEG data would be considered.

In the case of NeuCube, it was noticed how impractical an optimization process could be with all the samples (even that were just 543 in this work). However, a more in-depth analysis of the signals is necessary to develop an encoder adequate for these data. Also, just the LIF neuron model was used in the NeuCube for this work. Therefore, using other models could be useful to select the best for overt and imagined speech.

Then, although the SSN was the most analyzed classifier in this work (because it provided the best scores), it is still missing classification experiments with encoded EEG samples as inputs, similarly as in the NeuCube. These experiments could be relevant since these neuron models are based on the biological neurons, which transmit and receive spike trains.

Finally, another Spatio-temporal classifiers or classifier's adaptations to deal with spatial and temporal EEG data are suggested to be explored. One example is the use of deep recurrent-convolutional neural networks as those in [41, 42, 43]. Particularly in [41], based on the channel's locations, the EEG signals are transformed into sequential RGB topographic images, in which each resulting matrix represents the energy of a particular sub-band. Thus, the convolutional network processes each image to pass it to the recurrent network, which extracts the temporal information if the sequences.

---

---

## Bibliography

---

- [1] Shunan Zhao and Frank Rudzicz. Classifying phonological categories in imagined and articulated speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 992–996. IEEE, 2015.
- [2] Shunan Zhao and Frank Rudzicz. Combining different modalities in classifying phonological categories. In *Machine Learning and Interpretation in Neuroimaging*, pages 40–48. Springer, 2013.
- [3] The kara one database. <http://www.cs.toronto.edu/~complingweb/data/karaOne/karaOne.html>. Last access: 2019-11-03.
- [4] Maryann D’Alessandro, Rosana Esteller, George Vachtsevanos, Arthur Hinson, Javier Echauz, and Brian Litt. Epileptic seizure prediction using hybrid feature selection over multiple intracranial eeg electrode contacts: a report of four patients. *IEEE transactions on biomedical engineering*, 50(5):603–615, 2003.
- [5] Michael X Cohen. Where does eeg come from and what does it mean? *Trends in neurosciences*, 40(4):208–218, 2017.
- [6] Bruce Denby, Tanja Schultz, Kiyoshi Honda, Thomas Hueber, Jim M Gilbert, and Jonathan S Brumberg. Silent speech interfaces. *Speech Communication*, 52(4):270–287, 2010.
- [7] Patrick Suppes, Zhong-Lin Lu, and Bing Han. Brain wave recognition of words. *Proceedings of the National Academy of Sciences*, 94(26):14965–14969, 1997.
- [8] Marek Wester. Unspoken speech speech recognition based on electroencephalography, 7 2006.
- [9] Anne Porbadnigk, Marek Wester, and Tanja Schultz Jan-p Calliess. Eeg-based speech recognition impact of temporal effects. 2009.
- [10] Michael D’Zmura, Siyi Deng, Tom Lappas, Samuel Thorpe, and Ramesh Srinivasan. Toward eeg sensing of imagined speech. In *International Conference on Human-Computer Interaction*, pages 40–48. Springer, 2009.
- [11] Charles S DaSalla, Hiroyuki Kambara, Yasuharu Koike, and Makoto Sato. Spatial filtering and single-trial classification of eeg during vowel speech imagery. In *Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology*, page 27. ACM, 2009.
- [12] AA Torres-García, CA Reyes-García, L Villaseñor-Pineda, and JM Ramírez-Cortés. Análisis de señales electroencefalográficas para la clasificación de habla imaginada. *Revista mexicana de ingeniería biomédica*, 34(1):23–39, 2013.
- [13] Alejandro A Torres-García, Carlos A Reyes-García, Luis Villaseñor-Pineda, and Gregorio García-Aguilar. Implementing a fuzzy inference system in a multi-objective eeg channel selection model for imagined speech classification. *Expert Systems with Applications*, 59:1–12, 2016.

- [14] Jesus Salvador Garcia Salinas. Bag of features for imagined speech classification in electroencephalograms. 2017.
- [15] Germán A Pressel Coretto, Iván E Gareis, and H Leonardo Rufiner. Open access database of eeg signals recorded during imagined speech. In *12th International Symposium on Medical Information Processing and Analysis*, volume 10160, page 1016002. International Society for Optics and Photonics, 2017.
- [16] Yazaed Zaid Guadarrama Rendón. Procesamiento de señales eeg para el reconocimiento de voz imaginada. Master's thesis, Centro de Investigación en Computación, Instituto Politécnico Nacional, 6 2018.
- [17] Brainliner speech imagery dataset. <http://www.brainliner.jp/data/brainliner/speechImageryDataset>. Last access: 2019-11-03.
- [18] Sinci database. <https://drive.google.com/file/d/0By7apHbIp8ENZVBLRFV1SFhzBHC/view>. Last access: 2019-11-03.
- [19] Charles S DaSalla, Hiroyuki Kambara, Makoto Sato, and Yasuharu Koike. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural networks*, 22(9):1334–1339, 2009.
- [20] Diego A Pizzagalli et al. Electroencephalography and high-density electrophysiological source localization. *Handbook of psychophysiology*, 3:56–84, 2007.
- [21] Roberto A Vázquez. Pattern recognition using spiking neurons and firing rates. In *Ibero-American Conference on Artificial Intelligence*, pages 423–432. Springer, 2010.
- [22] Roberto A Vázquez and Beatriz A Garro. Training spiking neurons by means of particle swarm optimization. In *International Conference in Swarm Intelligence*, pages 242–249. Springer, 2011.
- [23] Arnaud Delorme and Scott Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21, 2004.
- [24] Lawrence R Rabiner and Marvin R Sambur. An algorithm for determining the endpoints of isolated utterances. *Bell System Technical Journal*, 54(2):297–315, 1975.
- [25] Importance of feature scaling. [https://scikit-learn.org/stable/auto\\_examples/preprocessing/plot\\_scaling\\_importance.html#sphx-glr-auto-examples-preprocessing-plot-scaling-importance-py](https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html#sphx-glr-auto-examples-preprocessing-plot-scaling-importance-py). Last access: 2019-11-03.
- [26] Balint Petro, Nikola Kasabov, and Rita M Kiss. Selection and optimization of temporal spike encoding methods for spiking neural networks. *IEEE transactions on neural networks and learning systems*, 2019.
- [27] Nikola Kasabov, Nathan Matthew Scott, Enmei Tu, Stefan Marks, Neelava Sengupta, Elisa Capecci, Muhamini Othman, Maryam Gholami Dotorjeh, Norhanifah Murli, Reggio Hartono, et al. Evolving spatio-temporal data machines based on the neucube neuromorphic framework: design methodology and selected applications. *Neural Networks*, 78:1–14, 2016.
- [28] Benjamin Schrauwen and Jan Van Campenhout. Bsa, a fast and accurate spike train encoding scheme. In *Proceedings of the International Joint Conference on Neural Networks, 2003.*, volume 4, pages 2825–2830. IEEE, 2003.
- [29] Nikola K Kasabov. Neucube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks*, 52:62–76, 2014.

- [30] Nuttapod Nuntalid, Kshitij Dhoble, and Nikola Kasabov. Eeg classification with bsa spike encoding algorithm and evolving probabilistic spiking neural network. In *International conference on neural information processing*, pages 451–460. Springer, 2011.
- [31] Rainer Storn and Kenneth Price. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.
- [32] Jason Brownlee. What is the difference between test and validation datasets? <https://machinelearningmastery.com/difference-test-validation-datasets/>, Jul 2017. Last access: 2019-11-03.
- [33] George Forman and Martin Scholz. Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. *ACM SIGKDD Explorations Newsletter*, 12(1):49–57, 2010.
- [34] Eugene M Izhikevich. Which model to use for cortical spiking neurons? *IEEE transactions on neural networks*, 15(5):1063–1070, 2004.
- [35] Eugene M. Izhikevich. Which model to use for cortical spiking neurons? <https://www.izhikevich.org/publications/whichmod.htm#izhikevich>, 2004. Last access: 2019-11-03.
- [36] Eugene M Izhikevich. Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572, 2003.
- [37] Eugene M. Izhikevich. Simple model of spiking neurons. <https://www.izhikevich.org/publications/spikes.htm>, 2003. Last access: 2019-11-03.
- [38] Stéphanie Martin. *Understanding and Decoding Imagined Speech using Electrocorticographic Recordings in Humans*. PhD thesis, École Polytechnique Fédérale de Lausanne, 7 2017.
- [39] Scott Makeig, Anthony J Bell, Tzyy-Ping Jung, and Terrence J Sejnowski. Independent component analysis of electroencephalographic data. In *Advances in neural information processing systems*, pages 145–151, 1996.
- [40] Vincent J Samar, Ajit Bopardikar, Raghuveer Rao, and Kenneth Swartz. Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. *Brain and language*, 66(1):7–60, 1999.
- [41] Pouya Bashivan, Irina Rish, Mohammed Yeasin, and Noel Codella. Learning representations from eeg with deep recurrent-convolutional neural networks. *arXiv preprint arXiv:1511.06448*, 2015.
- [42] Chuanqi Tan, Fuchun Sun, Wenchang Zhang, Jianhua Chen, and Chunfang Liu. Multimodal classification with deep convolutional-recurrent neural networks for electroencephalography. In *International Conference on Neural Information Processing*, pages 767–776. Springer, 2017.
- [43] Mehdi Hajinorozi, Zijing Mao, Yuan-Pin Lin, and Yufei Huang. Deep transfer learning for cross-subject and cross-experiment prediction of image rapid serial visual presentation events from eeg data. In *International Conference on Augmented Cognition*, pages 45–55. Springer, 2017.