POLITECNICO DI TORINO

DIPARTIMENTO DI INGEGNERIA MECCANICA E AEROSPAZIALE
(DIMEAS)

Master's Degree in Biomedical Engineering

Error estimation of feature extraction algorithms and weighted classification method for vocal signals

Supervisors
Prof. Alessio Carullo
PhD. Alessio Atzori
Prof. Alberto Vallan

Candidate
255464 - Tania Crisafulli

Turin, 19th July 2021

'A mia madre, a mio padre e a tutta la mia famiglia'

Abstract

Parkinson disease is a neurodegenerative disorder characterized by a slow but progressive evolution. Even if it mainly involves the motor system, issues with the phonatory system have been noticed. The patient loses full control of the speech apparatus; are noticed uncontrolled repetitions, incorrect articulation of words and a weakening of the voice speech.

In recent years, non-invasive techniques based on speech signal processing have been developed for the purpose of early diagnosis and to monitor the effects of pharmacological and neuro stimulation therapies.

This thesis work can be considered a primary step of a study focused on improving the models under analysis by increasing the database of the monitored subjects and identifying repeatable patterns using a weighted classification algorithm based on the estimate error of the feature extraction algorithm.

To acquire the vocal signals, wearable devices have been used that are able to monitor the subjects not impairing their daily activities. The analyzed signals have been recorded during the repetition of 3 vowels produced by 57 healthy subjects (HE), 67 patients with vocal tract diseases (PA) and 45 parkinsonian patients (PD). From this dataset very unbalanced in terms of both age and gender, a reduced balanced dataset has been extracted, which includes 10 subjects from each class.

The aim of the first part of this work has been the processing of the vocal signal to extract parameters that allow to evaluate the stability in frequency (jitter) and amplitude (shimmer) of the sustained vowels and other parameters related to the signal quality, such as harmonic to noise ratio and cepstral peak prominence smoothed (CPPS).

To increase the balanced dataset, artificial vowels with known sequences of periods and amplitudes have been generated through a Monte Carlo simulation with the Metropolis-Hastings algorithm.
The parameters extracted from the artificial signals have been compared to those

extracted from the original signals in order to estimate the error of the feature extraction algorithm.

The second part of the work has been focused on the evaluation of reproducibility and repeatability of the obtained measurements. The synthetic vowels previously obtained have been reproduced using a "Head and torso simulator" in an anechoic chamber. The vocal signals produced have been acquired using 3 different measurement chains:
- a microphone in air placed in 4 different positions;
- a reference phono-meter;
- a microphone embedded in an iPhone 8.

The same parameters have been extracted to evaluate the errors of the feature extraction algorithm by comparing the sequences obtained from these recordings to those extracted from the artificial signals.

In the third part of the work, the extracted features have been used to train a weighted logistic regression model to discriminate HE and PA subjects from PD subjects. The combination of features considered by the classifier were the one that had the lowest average relative error. The weights of the features of the algorithm are the reciprocal of the errors obtained in the previous steps.

The classification method, considering the original vowels, provided the probability of belonging to HE class with an accuracy of 84.2% and to PA class with an accuracy of 90.0%.
This method, using the weights obtained in the previous steps, provided the probability to belong to HE class with an accuracy 93.3% and to PA class with an accuracy of 93.3%.

Contents

List of Tables

List of Figures

Chapter 1

Parkison Disease

1.1 Characteristics of the Parkinson Disease

Parkinson's disease (PD) is a progressive, multi-systemic neurodegenerative disorder with motor and non-motor symptoms. The cardinal symptoms of the disease are bradykinesia (slowness and fatigue of movements), muscle stiffness and tremor. While bradykinesia and rigidity are invariably present (even if absent at the onset, they emerge in the evolution of the disease), tremor occurs only in a part of the patients [6]. These disorders can be associated with abnormalities of walking, posture and balance, which can appear relatively early in the progression of the disease.
Its prevalence in industrialized countries is about 0.3% (3 patients out of 1000) in the general population and increases up to 1% in people older than 60 and up to 4% of the population older than 80 [7].

1.2 Causes of the Disease

In Parkinson's disease, synuclein, that is a brain protein that helps nerve cells communicate, forms clusters called Lewy bodies that consists of misfolded synuclein. Synuclein can accumulate in different regions of the brain, in particular in the substantia nigra located deep inside the brain as shown in Figure 1.1 and interfere with brain function.

This leads to a degeneration of neurons in this area of the brain that causes the drop of the dopamine production, with a gradual progression and a prolonged course of the reduction in dopamine levels as shown in Figure 1.2. The causes are not yet known. Multiple elements appear to contribute to its development. These factors are mainly genetics, in fact some known mutations are associated with Parkinson's disease [8].

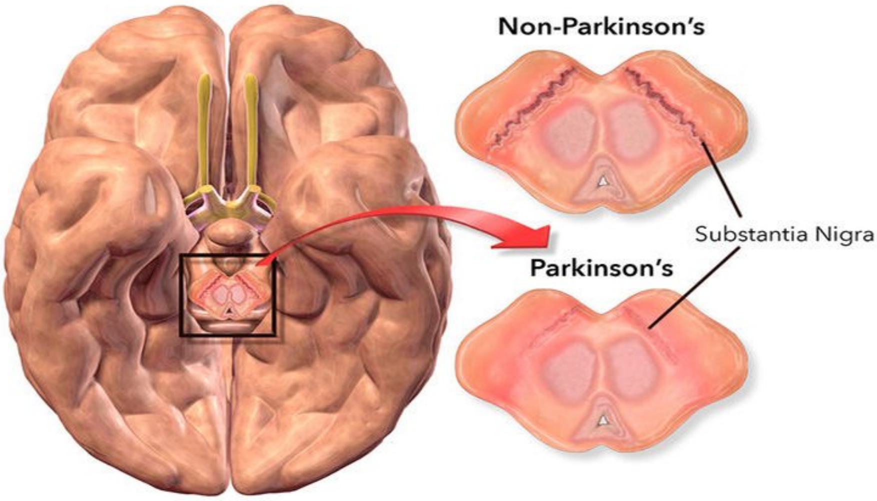

Figure 1.1: Substantia nigra differences between Parkinson's and non Parkinson's [1].

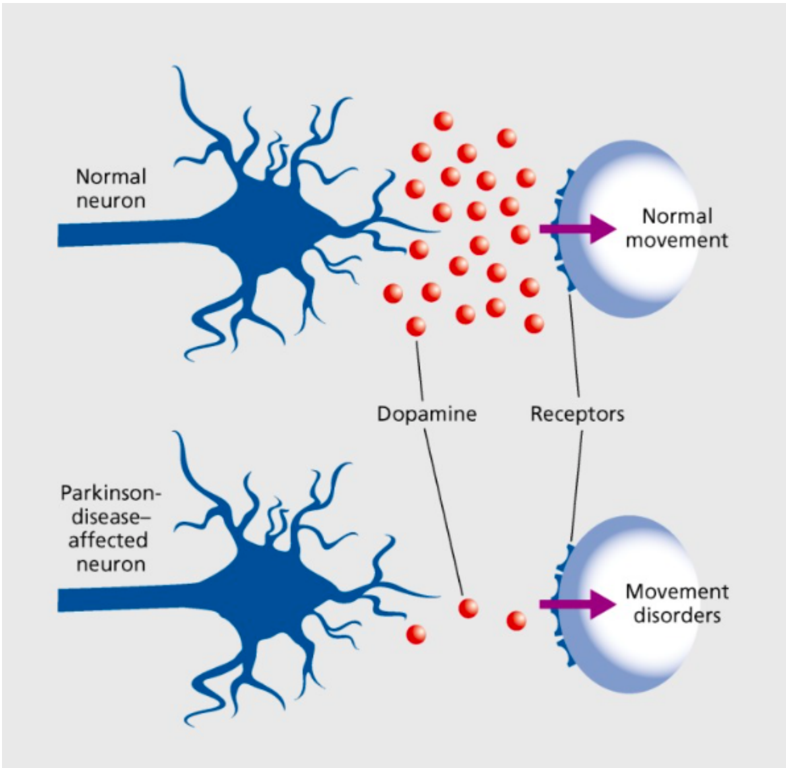

Figure 1.2: Different levels of dopamine[2].

About 20% of patients have a positive family history for the disease. It is estimated that family members of people with Parkinson's disease have a slightly

higher risk of developing the disease than the general population.
Among the genes identified the most important are: alpha-synuclein (PARK 1 / PARK 4), parkin (PARK-2), PINK1 (PARK-6), DJ-1 (PARK-7), LRRK2 (PARK-8) and GBA glucocerebrosidase.
There are also other factors, such as toxic factors or occupational exposure: the risk of illness acquires with toxins such as some pesticides or hydrocarbon-solvents and in some professions, the exposition of workers to heavy metals (iron, zinc, copper) [9].

Smoking appears to be a protective factor because smoke probably reduces the onset of Parkinson's disease.

Although movement disorders are the most obvious element of Parkinson's disease, the disorder can also impair other functions of the central nervous system such as cognitive processes, behavior, mood, night rest and the central nervous system. The alteration of one of these functions, or the parallel involvement of several systems can lead to the onset of so-called non-motor symptoms [9].

1.3 Diagnosis, disease assessment and therapies

To evaluate the severity of Parkinson's symptoms, although there are no specific diagnostic tests, some scales have been developed according to the Parkinson's Disease Society Brain Bank Diagnostic Criteria.
The most common are the Hoehn and Yahr scale and the UPDRS scale.

The Hoehn and Yahr Scale

The Hoehn and Yahr Scale measures the progression of Parkinson symtomps and the level of disability.
Originally published in 1967 in the journal Neurology by Melvin Yahr and Margaret Hoehn, it included stages 0 to 5 [10].

- Stage 0 - No signs of disease.
- Stage 1 - Symptoms on one side only (unilateral).
- Stage 1.5 - Symptoms unilateral and also involving the neck and spine.
- Stage 2 - Symptoms on both sides but no impairment of balance.
- Stage 2.5 - Mild symptoms on both sides.

- Stage 3 - Balance impairment, mild to moderate disease, physically independent.

- Stage 4 - Severe disability, but still able to walk or stand unassisted.

- Stage 5 - Needing a wheelchair or bedridden unless assisted.

The UPDRS Scale

The UPDRS scale (Unified Parkinson's disease rating scale), developed in 1987, is a very detailed system for evaluating the severity of Parkinson's symptoms. It is mostly by doctors to track the progression of patients' symptoms during the assumption of a particular medicine. UPDRS is characterized by four parts [11].

Part I : Cognitive skills, behavior and mood.

Part II: Daily activities.

Part III: Motor activity.

Part IV: Motor complications of therapy.

Thanks to this scale, is possible to obtain a numerical evaluation useful to compare the patient's results over time and follow the evolution of the disease.
A score equal to 0 represents complete disability and a vegetative level, while higher scores indicate a higher level of patient's ability and independence in performing daily tasks without difficulty.
Figure 1.3 shows the score sheet of the UPDRS scale.
Particular attention is payed for section 1 (Part III) that can be used as a "global" measure of communication assessment.
It includes diction, volume, intelligibility and expression, parameters which are used as index of the severity of the vocal system that establish the degree of disease progression.

________________________	___ ___ ___	___ ___ - ___ ___ - ___ ___ ___ ___ (mm-dd-yyyy)	___ ___ ___
Patient Name or Subject ID	Site ID	Assessment Date	Investigator's Initials

MDS UPDRS Score Sheet

1.A	Source of information	☐ Patient ☐ Caregiver ☐ Patient + Caregiver	3.3b	Rigidity– RUE	
			3.3c	Rigidity– LUE	
Part I			3.3d	Rigidity– RLE	
1.1	Cognitive impairment		3.3e	Rigidity– LLE	
1.2	Hallucinations and psychosis		3.4a	Finger tapping– Right hand	
1.3	Depressed mood		3.4b	Finger tapping– Left hand	
1.4	Anxious mood		3.5a	Hand movements– Right hand	
1.5	Apathy		3.5b	Hand movements– Left hand	
1.6	Features of DDS		3.6a	Pronation- supination movements– Right hand	
1.6a	Who is filling out questionnaire	☐ Patient ☐ Caregiver ☐ Patient + Caregiver	3.6b	Pronation- supination movements– Left hand	
			3.7a	Toe tapping–Right foot	
1.7	Sleep problems		3.7b	Toe tapping– Left foot	
1.8	Daytime sleepiness		3.8a	Leg agility– Right leg	
1.9	Pain and other sensations		3.8b	Leg agility– Left leg	
1.10	Urinary problems		3.9	Arising from chair	
1.11	Constipation problems		3.10	Gait	
1.12	Light headedness on standing		3.11	Freezing of gait	
1.13	Fatigue		3.12	Postural stability	
Part II			3.13	Posture	
2.1	Speech		3.14	Global spontaneity of movement	
2.2	Saliva and drooling		3.15a	Postural tremor– Right hand	
2.3	Chewing and swallowing		3.15b	Postural tremor– Left hand	
2.4	Eating tasks		3.16a	Kinetic tremor– Right hand	
2.5	Dressing		3.16b	Kinetic tremor– Left hand	
2.6	Hygiene		3.17a	Rest tremor amplitude– RUE	
2.7	Handwriting		3.17b	Rest tremor amplitude– LUE	
2.8	Doing hobbies and other activities		3.17c	Rest tremor amplitude– RLE	
2.9	Turning in bed		3.17d	Rest tremor amplitude– LLE	
2.10	Tremor		3.17e	Rest tremor amplitude– Lip/jaw	
2.11	Getting out of bed		3.18	Constancy of rest	
2.12	Walking and balance			Were dyskinesias presen	☐ No ☐ Yes
2.13	Freezing			Did these movements interfere with ratings?	☐ No ☐ Yes
3a	Is the patient on medication?	☐ No ☐ Yes		Hoehn and Yahr Stage	
3b	Patient's clinical state	☐ Off ☐ On	**Part IV**		
3c	Is the patient on Levodopa?	☐ No ☐ Yes	4.1	Time spent with dyskinesias	
3.C1	If yes, minutes since last dose:		4.2	Functional impact of dyskinesias	
Part III			4.3	Time spent in the OFF state	
3.1	Speech		4.4	Functional impact of fluctuations	
3.2	Facial expression		4.5	Complexity of motor fluctuations	
3.3a	Rigidity– Neck		4.6	Painful OFF-state dystonia	

Figure 1.3: Patients evaluation: score sheet - UPDRS scale [3].

1.4 The affection of the vocal system

In Parkinson's disease, apart from motor symptoms there are also non motor symptoms that are not often considered.
However, speaking difficulties arise in about half of the affected patients while the other half of the patients, even after many years from the onset of the disease, may not experience such problem. Changes in the ability to communicate can result in the tendency to self social isolation.
Speaking is a complex motor task which implies the recruitment of some muscles and in particular those that control breathing, phonation (emission of the voice), articulation (pronunciation), prosody (rhythm, intonation and pace of speech).
In Parkinson's disease the alterations of the voice are due to a reduced coordination of these muscles.

The symptoms that can occur as a consequence of these alterations are:

- Weakening of the volume of the voice

It is the first change to be noticed. Over time, this reduction can come to the point of inaudibility of the voice.

- Dulling of the voice

The voice is strong at the beginning of a sentence, but fades at the and of any sentence vocalization.

- Monotone voice

The voice remains at the same level, does not vary and lacks expression.

- Voice quality change

The sound of the voice is trembling, faint or more acute.

- Involuntary hesitation before speaking

The difficulty to start talking and keep the voice steady from beginning to end of a conversation or a phrase.

- Indistinct articulation (pronunciation)

The pronunciation in particular the end of the words is omitted; the sounds of the final syllables are unclear

- Fast accelerated pace

The syllables and the words flow without pauses.
There could be also a progressive acceleration of words towards the end of a sentence.

- Uncontrolled repetitions

Words, phrases and sentences are repeated involuntarily and out of control [12].

Chapter 2

State of the Art

2.1 Introduction

The voice in normal speech production is a non stationary signal.
The characteristics of the signal vary quickly over time and change accordingly to the particular sound emitted.
Such signal, in speech and reading, can be considered, with good approximation, stationary during a window of about 10-30 ms.
It is characterized by a fundamental period T_0, a fundamental frequency that is the inverse of the period $F_0 = \frac{1}{T_0}$ and by its harmonics that have frequencies that are integer multiples of the fundamental frequency. The value of F_0 oscillate around an average value, which is a characteristic of each individual and varies according to age, gender and type of vocal activity as shown in Table 2.1:

Frequency	**Type**
105 ÷ 160 Hz	Men
255 ÷ 440 Hz	Children
175 ÷ 245 Hz	Women [13]

Table 2.1: Frequency range depending on gender and age

The vocal apparatus can be considered as an exciter-resonator system. The vocal folds represent the exciter and the vocal tract is the resonator. The modification of the shape of the vocal tract is involved in the articulation of phonemes in speech production.

However, the position of the fundamental frequency F_0 remains unchanged, only the amplitudes of the various spectral components are modified. The audible spectrum can have components up to 20 kHz, but most of the energy is concentrated below 4 kHz in speech production.
The scientific and technological progress of the last decades has contributed to an effective revolution in the study and analysis of the vocal signal. Modern acoustic analysis is in fact based on the use of computers with dedicated hardware and software.

2.1.1 Acoustic parameters

In the past years various researches focused on the correlation between objective parameters extracted from the vocal signals and several pathologies that can influence the voice production.
Considering the production of continuous vowels, the analysis is focused on the characteristics of signals in the time and frequency domain. Particular attention can be payed to the relative perturbations of such measurement and to the relationship between harmonic and non-harmonic components [14].
In this work, the acoustic parameters used to describe signals perturbations to evaluate the voice quality are:
• Local Jitter (%): Relative measure of the variation of the fundamental pseudo-period from cycle to cycle.

$$Jitt = 100\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_0^{(i)} - T_0^{(i+1)}|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \tag{2.1}$$

where $T_0^{(i)}$, i = 1,2...N are the estimated pseudo-periods, N is the number of estimated pseudo-periods.

• Absolute jitter (μs): absolute measure of the variation of the fundamental pseudo-period cycle to cycle.

$$Jita = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_0^{(i)} - T_0^{(i+1)}| \tag{2.2}$$

where $T_0^{(i)}$, i = 1,2...N are the estimated pseudo-periods, N is the number of estimated pseudo-periods.

• RAP (%): relative measure of the change in pitch from period to period with an average over 3 successive periods.

$$RAP = 100 \frac{\frac{1}{N-2}\sum_{i=2}^{N-1} |\frac{T_0^{(i-1)} - T_0^{(i)} - T_0^{(i+1)}}{3} - T_0^{(i)}|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \tag{2.3}$$

where $T_0^{(i)}$, i = 1,2...N are the estimated pseudo-periods, N is the number of estimated pseudo-periods.

- PPQ (%): relative measure of the pitch variation from period to period with an average over 5 successive periods.

$$PPQ = 100 \frac{\frac{1}{N-4}\sum_{i=1}^{N-4} |\frac{1}{5}\sum_{r=o}^{4} T_0^{(i+r)} - T_0^{(i+2)}|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \tag{2.4}$$

where $T_0^{(i)}$, i = 1,2...N are the estimated pseudo-periods, N is the number of estimated pseudo-periods.

- vF_0 (%): represents the relative standard deviation of the frequency basic.

$$vF_0(\%) = 100 \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N} (F_0^{(i)} - \overline{F})^2}}{\overline{F}} \tag{2.5}$$

where $F_0^{(i)} = \frac{1}{T_0^{(i)}}$ are the values of the fundamental frequencies of each estimated pseudo-period, $\overline{F}$ is the mean value of $F_0^{(i)}$, $T_0^{(i)}$, i = 1,2...N are the estimated pseudo-periods, N is the number of estimated pseudo-periods.

- Shimmer percentage (%): relative measure of the amplitude variation peak-to-peak period to period.

$$Shim = 100 \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} |A_0^{(i)} - A_0^{(i+1)}|}{\frac{1}{N}\sum_{i=1}^{N} A_0^{(i)}} \tag{2.6}$$

where $A^{(i)}$, i = 1,2...N are the peak-to-peak amplitudes, N is the number of estimated pulses.

• Absolute shimmer (dB): measurement of the variation of the peak amplitude - peak from period to period. It is calculated as:

$$ShidB = \frac{1}{N-1}\sum_{i=1}^{N-1}|20log(A_0^{(i)}/A_0^{(i+1)})| \tag{2.7}$$

where $A^{(i)}$, i = 1,2...N are the peak-to-peak amplitudes, N is the number of estimated pulses.

• APQ (%): measure of the variation of the peak-to-peak amplitude from period to period with an average over 11 successive periods. It is calculated as:

$$APQ = 100\frac{\frac{1}{N-10}\sum\limits_{i=1}^{N-10}|\frac{1}{11}\sum\limits_{r=o}^{10}A_0^{(i+r)} - A_0^{(i+5)}|}{\frac{1}{N}\sum\limits_{i=1}^{N}A_0^{(i)}} \tag{2.8}$$

where $A^{(i)}$, i = 1,2...N are the peak-to-peak amplitudes, N is the number of estimated pulses.
• vAm (%): relative standard deviation of the peak-to-peak amplitude. It is calculated as:

$$vAm(\%) = 100\frac{\sqrt{\frac{1}{N}\sum\limits_{i=1}^{N}(A_0^{(i)} - \overline{A})^2}}{\overline{A}} \tag{2.9}$$

where $A_0^{(i)}$, i = 1,2...N are the peak-to-peak amplitudes, $\overline{A}$ is the mean value of $A_0^{(i)}$, N is the number of estimated pulses.

Another extracted parameter, is the Harmonics-to-Noise Ratio. This parameter is calculated as the ratio between the harmonic spectral energy components and the unharmonic spectral energy components, evaluating the ratio between the harmonic and non harmonic content of the speech signal.
The formula for calculating the HNR, based on autocorrelation is [15]:

$$HNR = 10log\frac{AC_v(\tau)}{AC_v(\tau) - AC_v(0)} \tag{2.10}$$

where $AC_v(0)$ is the value of the autocorrelation function considering a delay equal to 0 (the power of the signal) while $AC_v(\tau)$ is the value of the autocorrelation function when the delay is equal to τ, which is the estimated pseudo-period [15].

Under this assumption, the HNR assumes a similar physical meaning of the Signal-to-Noise Ratio (SNR). Higher HNR values indicate a periodic component in the signal greater than the noise component while lower values of HNR indicate a higher noise component than the periodic component. Ideally periodic signal present an infinite HNR.

Another parameter based on autocorrelation is the fundamental frequency F_0, which is defined as the inverse of the pseudo-period T_0. This parameter is particularly important because is used to calculate stability parameters as seen in equations 2.1 - 2.9.

The signal frames are labeled according to their magnitude and harmonic content. In particular, three kinds of frame have been labeled:
• silent frames;
• unvoiced frames (containing unvoiced sounds);
• voiced frames (containing vocalized sounds).

The frame labeling has been performed using the following algorithm:

1. The signal is sliced into frames:

• if $RMS_{frame} < \frac{RMS_{signal}}{2}$ where RMS_{signal} is the root mean square of the entire signal, then the interval considered is a frame of silence;

• otherwise go to the next step.

2. The HNR and the fundamental frequency F_0 is evaluated:

• if $[\frac{F_{0i}-F_{0i-1}}{F_{0i-1}}] < 0.5$ and $[HNR_i > 0$ dB] is false, then the frame is "Unvoiced";

• otherwise the frame is "Voiced".

According to this algorithm, is possible to calculate another parameter of interest which is the ratio between vocalized sounds and unvoiced sounds vocalized:

• V/uV $= 100\frac{(frames)_v}{(frames)_{uv}}(\%)$.

The validity of these parameters is limited by the difficulty in determining the fundamental frequency, in fact small errors can become relevant in the measurement of perturbations. Besides, it also changes with gender and age, thought to depend on factors such as the state of mind of the person, voice education, voice professional use and lifestyle. These measurements also, in fact are valid only on vocal material consisting of sustained vowels.

2.1.2 CPP and CPPS

In recent years, acoustic parameters that are not based on the estimation of the fundamental frequency have been developed. Such evaluations can also be adapted to free speech [16].

Two important parameters are the Cepstral Peak Prominence (CPP) and the Cepstral Peak Prominence Smoothed (CPPS).

The concept of cepstrum was introduced in 1963 by Bogert et al.
Is defined as “the logarithmic power spectrum of the logarithmic spectrum of the speech signal” [17]. For this reason, the terms “spectrum” and “frequency” can be replaced by the corresponding words “cepstrum” and “quefrency”.

Cepstrum is defined as :

$$C(\tau) = |\mathcal{F}\{log(|\mathcal{F}\{f(t)\}|^2)\}|^2 \tag{2.11}$$

where $\mathcal{F}$ is the Fourier transform, $|\mathcal{F}\{f(t)\}|^2$ is the power spectrum and $f(t)$ is the signal as a function of time.

The variable τ is the "quefrency" mentioned before and has the dimensions of time.

CPP is an acoustic measure of the degree of harmony within a voice sample.
The more periodic the voice signal, the greater is harmonicity and the greater is the value of the CPP [18].

Unlike other acoustic measures that rely on pitch tracking mechanisms to measure the degree of perturbation within a voice signal, CPP is based on spectral transformation of fixed length frames [18].

The CPP is the measure in dB of the amplitude prominence of the first cepstral peak (also called as first “rahmonic”) measured as the distance from the regression line of the cepstrum floor.
Harmonic signals show a prominent cepstral peak at the index corresponding to the fundamental period, whereas dysphonic voice signals with disturbed periodicity are

associated with a reduced prominence of cepstral peak.

To correctly estimate the cepstral peak prominence it is necessary to search it in a quefrency range from 3 ms to 16 ms, which corresponds to the frequency range from 60 Hz to 300 Hz because in this range is concentrated the F_0 of the human voice.

To calculate CPPS, measured in dB, two smoothing steps are needed [19].

In the first phase the cepstra are time averaged: the current cepstrum is averaged with a certain number of cepstra preceding and following the considered cepstrum.

In second phase the cepstra are averaged along the quefrency.

In this way a CPPS is defined for each single frame and the output of the algorithm is composed of different values of CPPS.
From the evaluated CPPS, vectors statistical distribution can be extracted to evaluate statistical values such as mean, median, mode, standard deviation, range, 5° percentile, 95° percentile, skewness and kurtosis.

Chapter 3

Materials and Methods

3.1 Signal acquisition

3.1.1 Data collection

The data provided have been collected at "Città della Salute" in Turin. The participants who took part in the research were:

- 57 healthy subjects (HE).
- 67 patients with pathological non-Parkinson's disease (PA).
- 45 patients with Parkinson's disease (PD).

Both HE subjects and PA subjects underwent an endoscopic analysis to certify the subjects' vocal apparatus health status. The Figure 3.1 shows the dataset provided.
The dataset is unbalanced both in terms of gender and age (Figure 3.2) and (Figure 3.3) and in terms of the number of the subjects (Figure 3.4).

	HE	PD	PA
Total number of Patients	57	45	67
Number of Men	30	27	26
Number of Women	27	18	41
Average Age	30.7	66.8	52.5
Standard Deviation (Age)	1.7	1.9	2.3

Figure 3.1: Unbalanced Dataset considering gender and average age of patients.

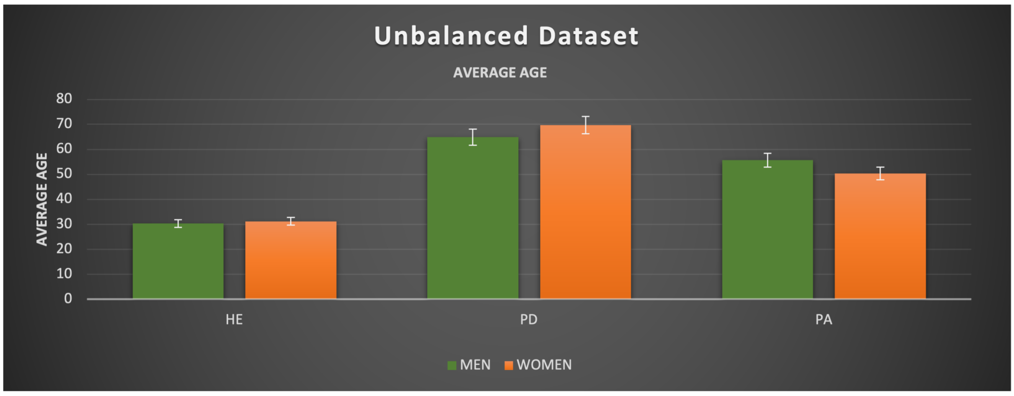

Figure 3.2: Unbalanced Dataset considering gender and average age of patients.

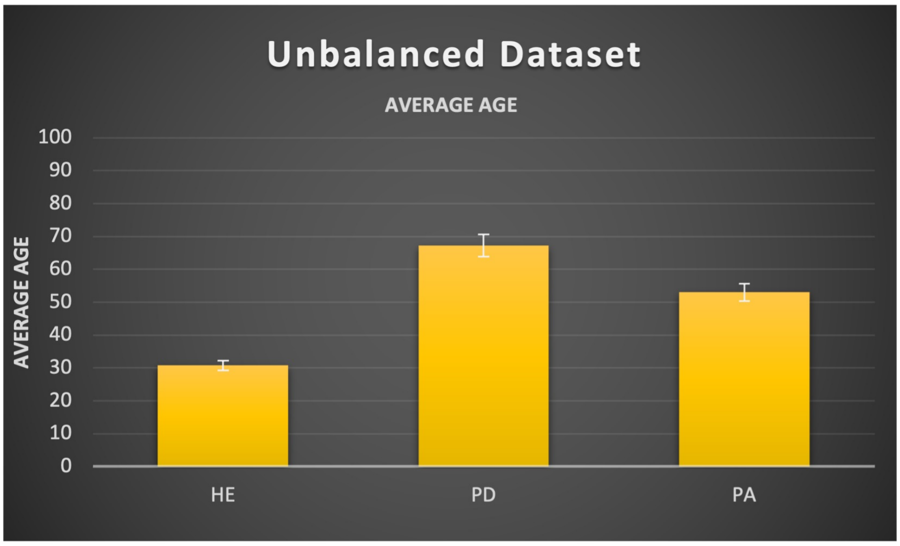

Figure 3.3: Unbalanced Dataset considering averag age of patients.

To overcome the problem with the unbalanced set of data collected, a balanced set of data has been obtained considering 5 males and 5 females of healthy, pathological and parkinsonian patients.
As shown in Figure 3.5, the provided dataset is balanced both in terms of gender and age (Figure 3.6) and (Figure 3.7) and in terms of number of subjects (Figure 3.8).

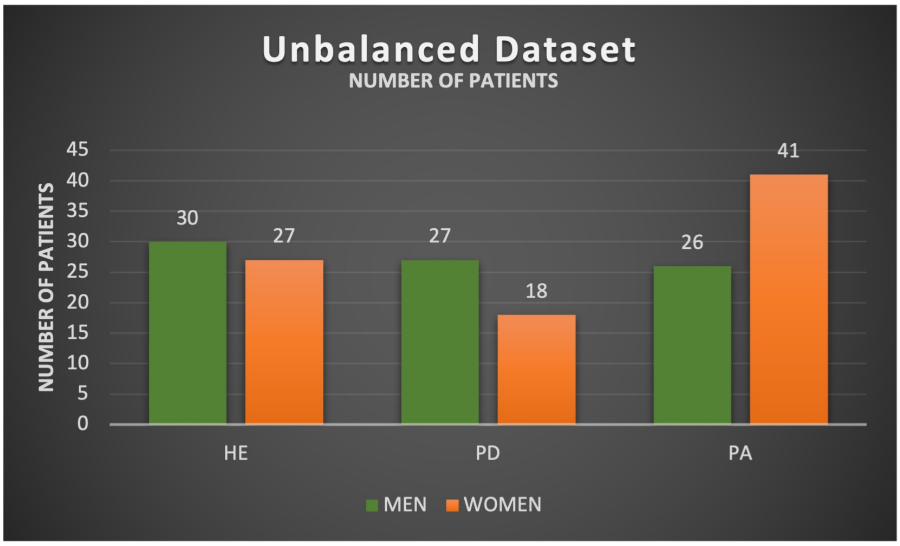

Figure 3.4: Unbalanced Dataset considering number of patients.

	HE	PD	PA
Total number of Patients	10	10	10
Number of Men	5	5	5
Number of Women	5	5	5
Average Age	52,95	54	52,8
Standard Deviation (Age)	1.77	1.78	1.61

Figure 3.5: Balanced Dataset considering gender and average age of patients

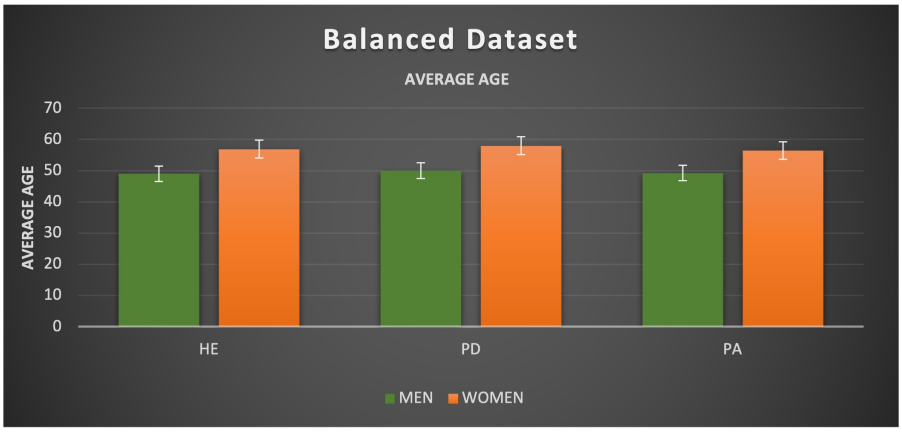

Figure 3.6: Balanced Dataset considering gender and average age of patients

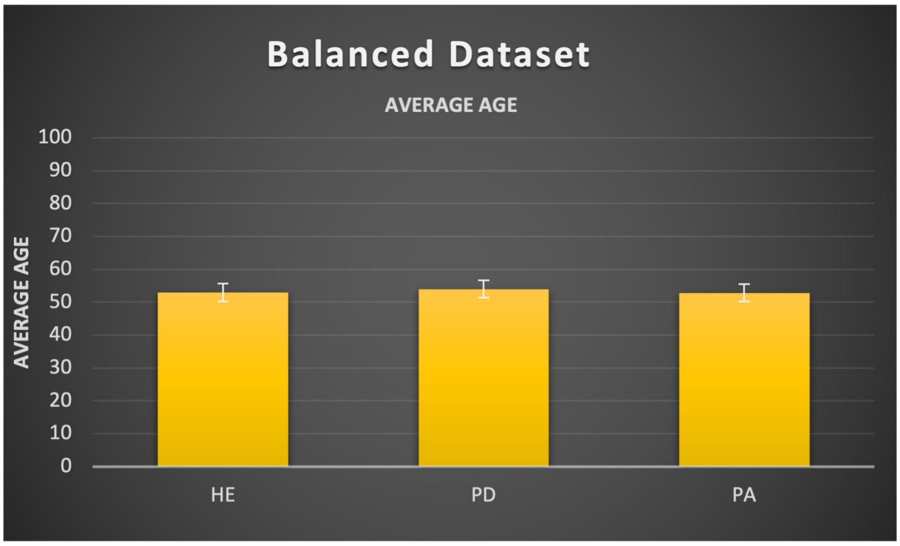

Figure 3.7: Balanced Dataset considering average age of patients.

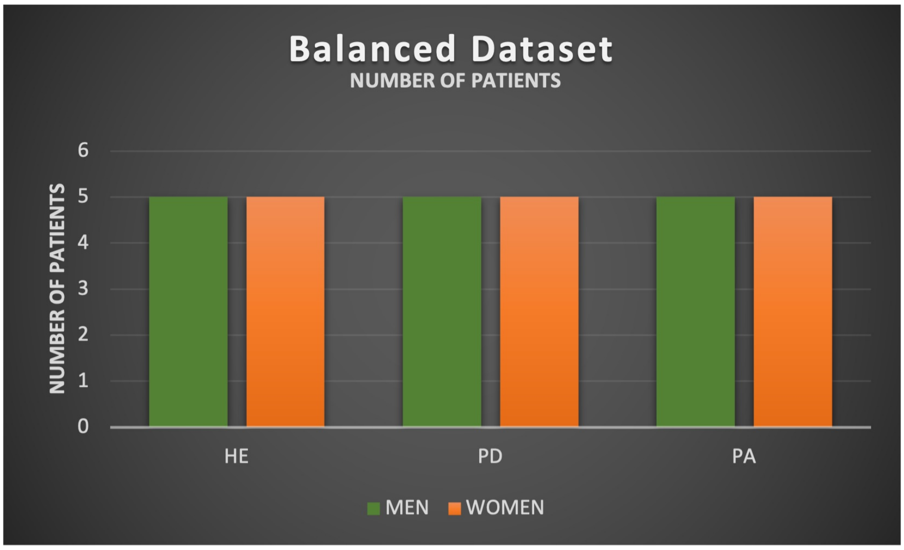

Figure 3.8: Balanced Dataset considering number of patients.

3.1.2 Data acquisition chain

For the voice signals acquisition of the subjects monitored in this study, two kind of microphones have been used:

- a cheek-type microphone in air (MIPRO MU55-HR) placed at a distance of about 2.5 cm from the subject's lips; an example is shown in Figure 3.9.

Figure 3.9: Microphone in Air

- a piezoelectric contact microphone (HX505-1-1), which is a collar whose sensitive element must be positioned near the jugular notch to pick up the vibrations of the vocal cords; an example is shown in Figure 3.10.

Figure 3.10: Contact Microphone

Both microphones have been connected to a portable recorder (EDIROL ROLAND R09-HR) using its stereo input.

The portable recorder is shown in Figure 3.9 and in Figure 3.10.

All recordings have been made with a sampling rate Fs = 44100 Sa/s and bit resolution of 16-bit.

The contact microphone is insensitive to the background noises so is suitable for noisy environments and for long term evaluations. However such microphone is insensitive to unvoiced phonemes such as "/s/" and "/f/" so the use of a microphone in air is needed to have a complete evaluation of voice signals.

Figure 3.11 shows a comparison between the same signal acquired using a microphone in air and a conctact microphone from a Pathological non Parkinsonian patient.

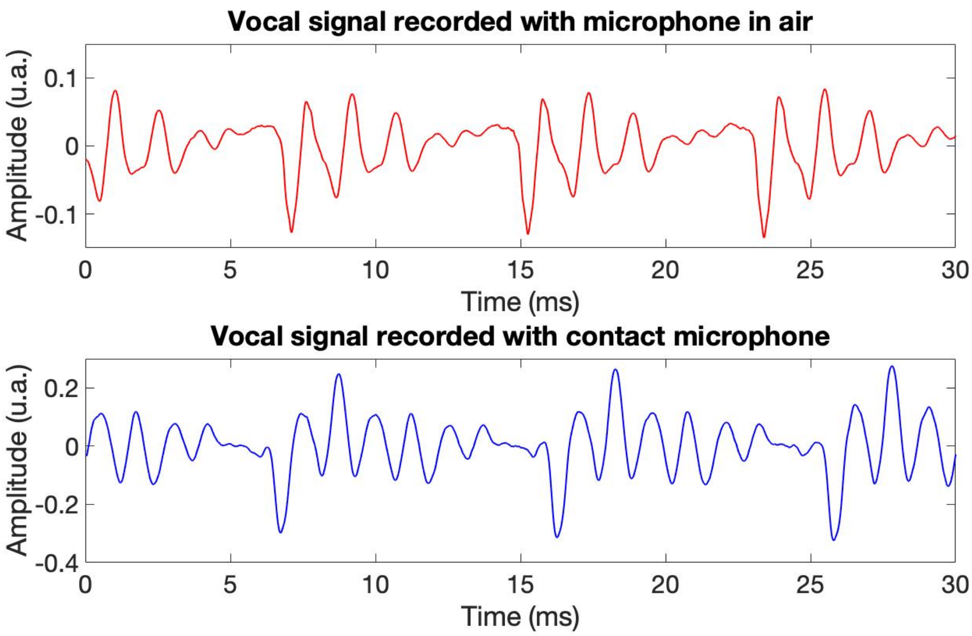

Figure 3.11: Signal recorded with both Microphone in air and Contact Microphone

In this thesis work, only recordings made by microphone in the air, have been considered.

3.1.3 Acquisition protocol and Pre-processing

The acquisitions of the vocal signals, required a well-defined protocol. This protocol was structured in 3 distinct parts:

• production of the '/a/' phoneme maintained for as long as possible ($3 \div 10$ seconds), at a confortable intensity.
In this case, the subjects involved were asked to repeat the voice emission three times.

• reading of a phonetically balanced text, reported in AppendixA, for about 1 minute;

• eloquence on a free topic lasting about 1 minute.

The recorded samples have been processed using the Audacity software:

• sustained vowel '/a/': for each recording, the three vowels were separated into individual files called A1, A2, A3, eliminating the initial and final part of the recording to avoid signal instability;

• reading: the recording of the reading was cut from the beginning of the piece 'Bulka' to the word 'sanguisuga', in order to reduce computational times and to balance the analysis time with that of free speech.

• free speech: the part relating to free speech has been cut from the recording from the beginning of speech up to a maximum of 1 minute to comply with the analysis duration of the reading file and reduce computational times.

In this thesis work, the attention has been focused on the analysis of the repetition of the sustained vowel recorded by the microphone in air.

3.2 Extracted parameters

The recorded samples have been processed in order to extract significant acoustic parameters of the three groups.
These parameters, for clarity sake, can be divided in different groups and referenced with a number used in the following chapters.

Time-stability parameters:

Parameter	Reference
Jita	1
Jitt	2
RAP	3
PPQ	4
vF0	5

Table 3.1: Frequency parameters

Amplitude-stability parameters:

Parameter	Reference
Shim	6
ShdB	7
APQ	8
VAm	9

Table 3.2: Amplitude parameters

HNR temporal evolutions are quantized considering the harmonic frames to get statistical distributions and calculate some relevant parameters. Table 3.3 shows the extracted parameters.

Parameter (HNR)	Reference
Mean	10
Median	11
Mode	12
Standard Deviation	13
Range	14
5° Percentile	15
95° Percentile	16
Skewness	17
Kurtosis	18

Table 3.3: HNR parameters

F_0 temporal evolutions are quantized considering the harmonic frames to get statistical distributions and calculate some relevant parameters. Table 3.4 shows the extracted parameters.

Parameter ($\mathbf{F_0}$)	**Reference**
Mean	19
Median	20
Mode	21
Standard Deviation	22
Range	23
5° Percentile	24
95° Percentile	25
Skewness	26
Kurtosis	27

Table 3.4: F_0 parameters

RMS temporal evolutions are quantized considering the harmonic frames to get statistical distributions and calculate some relevant parameters. Table 3.5 shows the extracted parameters.

Parameter(RMS)	**Reference**
Mean	28
Median	29
Mode	30
Standard Deviation	31
Range	32
5° Percentile	33
95° Percentile	34
Skewness	35
Kurtosis	36

Table 3.5: RMS parameters

CPPS temporal evolutions are quantized considering the harmonic frames to get statistical distributions and calculate some relevant parameters. Table 3.6 shows the extracted parameters.

Parameter(CPPS)	**Reference**
Mean	37
Median	38
mode	39
Standard Deviation	40
Range	41
5° Percentile	42
95° Percentile	43
Skewness	44
Kurtosis	45

Table 3.6: CPPS parameters

The last parameter, shown in Table 3.7 is:

Parameter	Reference
V/uV	46

Table 3.7: Voiced/Unvoiced parameter

In Figure 3.12 an example of Jitter extracted from PD and HE subjects is shown. The results obtained show how Parkinsonian subjects have higher Jitter than healthy subjects.

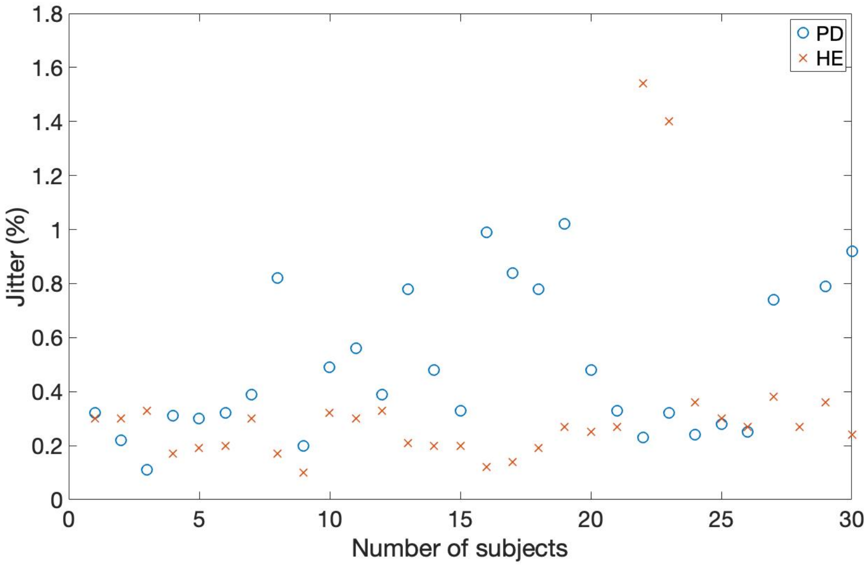

Figure 3.12: Jitter differences between PD and HE patients

In Figure 3.12 an example of Shimmer extracted from PD and HE subjects is shown. The results obtained show how Parkinsonian subjects have higher Shimmer than healthy subjects.

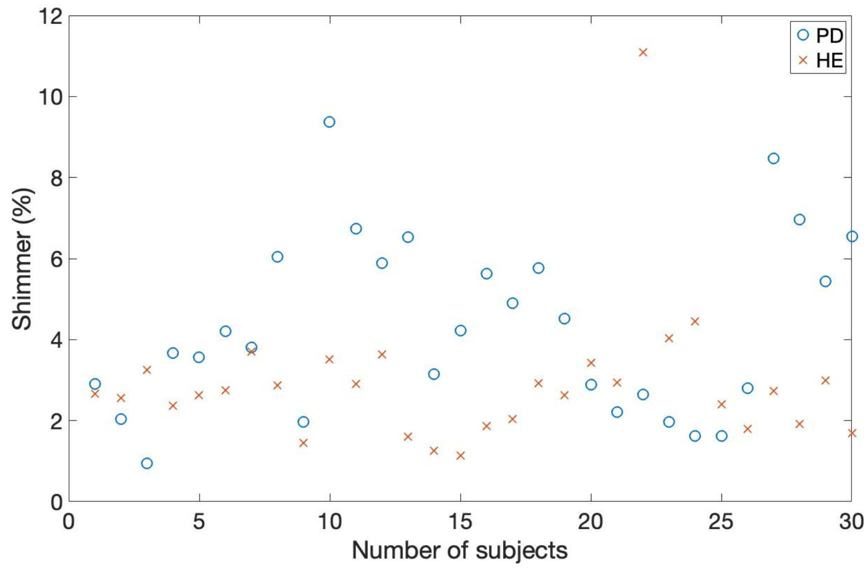

Figure 3.13: Shimmer differences between PD and HE patients

3.3 Generation of the artificial vowels

3.3.1 Introduction

This section describes the method used to generate artificial vowels using a Monte Carlo algorithm in order to increase the balanced dataset.

3.3.2 The MonteCarlo Method

The Monte Carlo (MC) method is an algorithm based on the generation of random variables starting from assigned probability distributions [20].
In this work the MC algorithm has been used to generate temporal sequences that have statistical distributions compatible with the reference statistical distributions. One of the most famous MC algorithms is the Metropolis-Hastings algorithm [21], which has been used to generate the artificial vowels using sequences of periods and amplitudes obtained from the feature extraction algorithm under test.

Such algorithm considers the sequence of periods extracted from a real vowel signal and its distribution D_t. From the periods and amplitudes vectors, a sequence of consecutive differences is obtained (es $\delta T = T_i - T_{i-1}$ for all i $\in$ [1,i$_{max}$]) and the correspondent distributions $D(\delta T)$ and $D(\delta A)$ are evaluated.
The empirical cumulative distributions are calculated using the matlab function "ecdf".
The generation algorithm is composed by five steps:

1. Through the direct MC method a proposal jump is generated as the inverse of the cumulative function using as input an uniform random number between 0 and 1.

2. The proposal jump is added to the period extracted from the actual signal, in order to obtain a new $T(i)$ value.

3. The proposal metric is calculated as:

$$A = min(1, \frac{D_t(i)}{D_t(i-1)}) \tag{3.1}$$

4. A random uniform number u between 0 and 1 is generated;

- the proposal is accepted if $u <= $ A;

- the proposal is rejected if $u >$ A.

5. The generation goes on until the Kolmogorov-Smirnof test, performed using the Matlab function "kstest2", confirms the compatibility between the statistical distributions of the real and the artificial signal with a confidence interval of 99%.

Figure 3.14 shows the distribution extracted from a vocal signal recorded by a female patient affected by Parkinson disease and the distribution that corresponds to an artificial signal.
The shown distributions are statistically compatible.

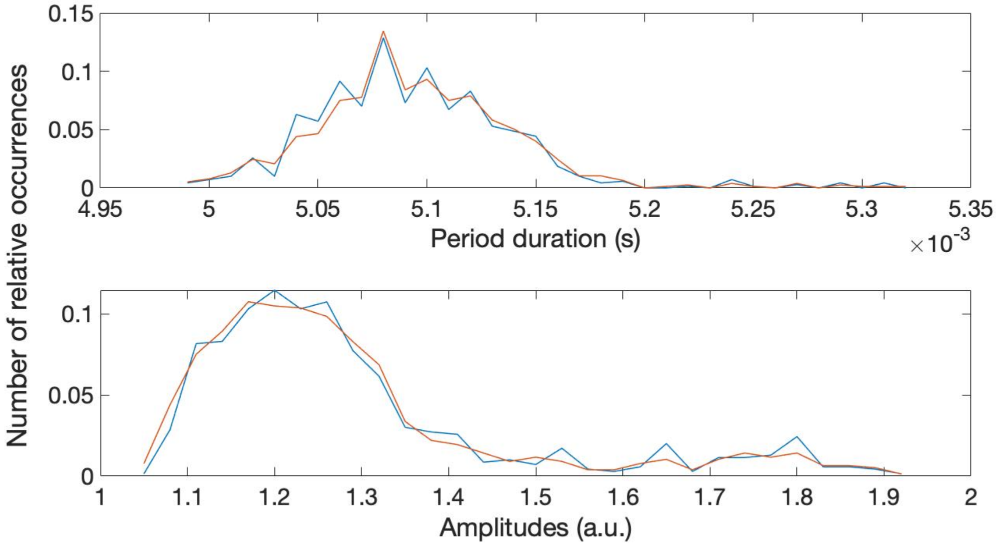

Figure 3.14: Periods and amplitudes distributions of the real vowel recorded by a Parkinsonian subject in orange and artificial vowel in blue.

3.3.3 From Numbers To Signals: the Concatenation Method

To transform the numeric sequences of periods and amplitudes generated by the MC algorithm in an artificial vowel signal the following method has been developed.

The real voice signal has been divided into pseudo-periods, the same used for the MC generation.

The modification of the length of the single periods took place through a linear resampling using the matlab function "linspace", while the modification of the amplitudes occurs through the equation:

$$y_1(t) = \frac{y_0(t)A_1}{A_0} \tag{3.2}$$

where $y_1(t)$ is signal vector of the desired $i - th$ pseudo-period candidate, $y_0(t)$ is the original pseudo-period vector, A_1 is the peak-to-peak amplitude value generated, A_0 is the value of the original peak-to-peak amplitude.

Lastly, the new frames have been concatenated one after the other, generating an artificial signal.

Figure 3.15 and 3.16 show a comparison between the original recorded signal and the obtained artificial signal highlighting the amplitude positive peaks.

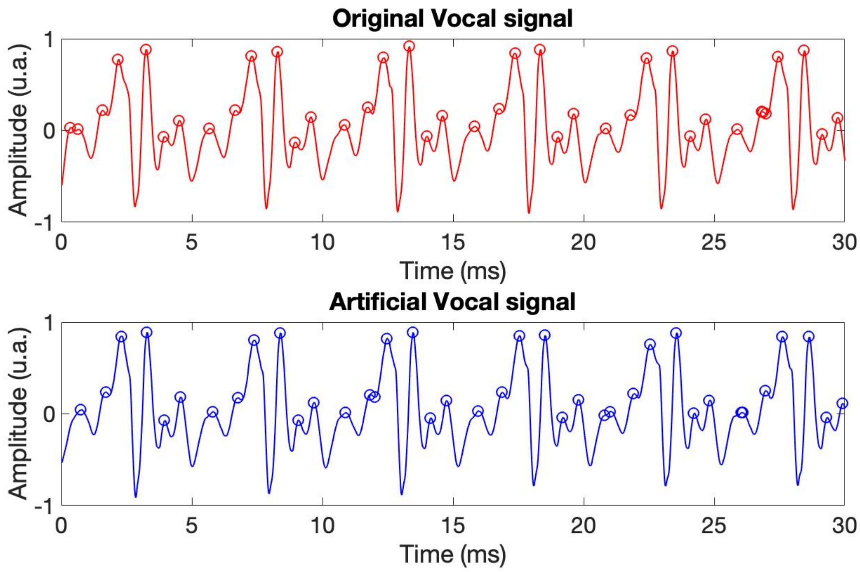

Figure 3.15: Comparison between the original signal and the artificial signal of a Parkinsonian subject plotted separately, highlighting the amplitude positive peaks.

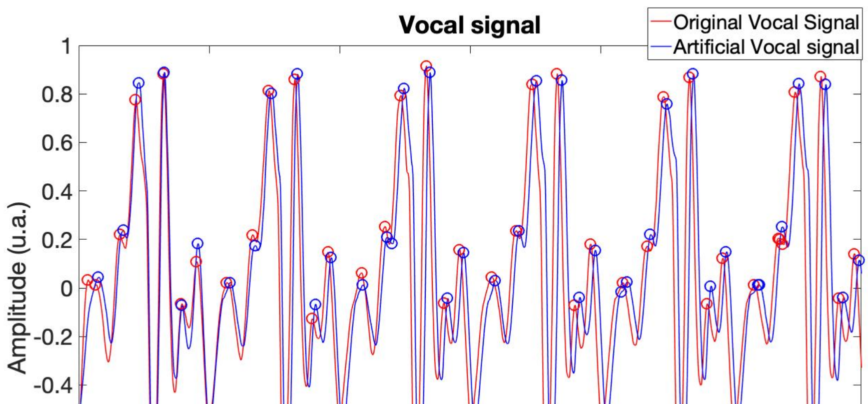

Figure 3.16: Comparison between the original signal and the artificial signal of a Parkinsonian subject highlighting the amplitude positive peaks.

3.4 Measurement reproducibility and repeteability

In the second part of the work, the attention has been focused on the reproducibility and repeatability of the performed measurements.
To achieve this, a ”Head and Torso Simulator” (HATS) has been used to emit artificial vowels in an anechoic room.

3.4.1 Head and Torso Simulator

The ”Head and Torso Simulator” (HATS), Type 4150-C is a test dummy with a mouth simulators and ear simulators [22]. The Type 4150-C simulator is characterized by a torso over which is mounted a head. The HATS, allow to represent the standardized average dimensions of a human adult, and offers a reproducible simulation of its acoustic parameters and properties [22].
The model used was fitted on a turntable using a chair. Figure 3.17 and Figure 3.18 show the HATS used in the anechoic chamber.

Figure 3.17: Pictures of HATS - frontal view in the anechoic chamber.

Figure 3.18: Pictures of HATS - lateral view in the anechoic chamber.

3.4.2 HATS Characteristics

The HATS mouth simulator has a high-compliance loudspeaker that gives powerful low-frequency response and low distortion.
The acoustic transmission path from the loudspeaker to the mouth opening ensures an easily equalized frequency response of the sound pressure level in front of the mouth [4].
The mouth simulator produces a sound-pressure distribution around the opening of the mouth simulating that of a median adult human mouth and it follows the frequency range of human subjects [22].

The loudspeaker has an impedance of 4 ohm and a maximum power rating for continuous operation of 10 W.
To reduce the risk of damage, the drive to the loudspeaker is limited by a protection circuit mounted in the head of Type 4150-C.
Figure 3.19, shows the sound-pressure level spatial distribution around the mouth obtained doing a measurement with the mouth insert in place.

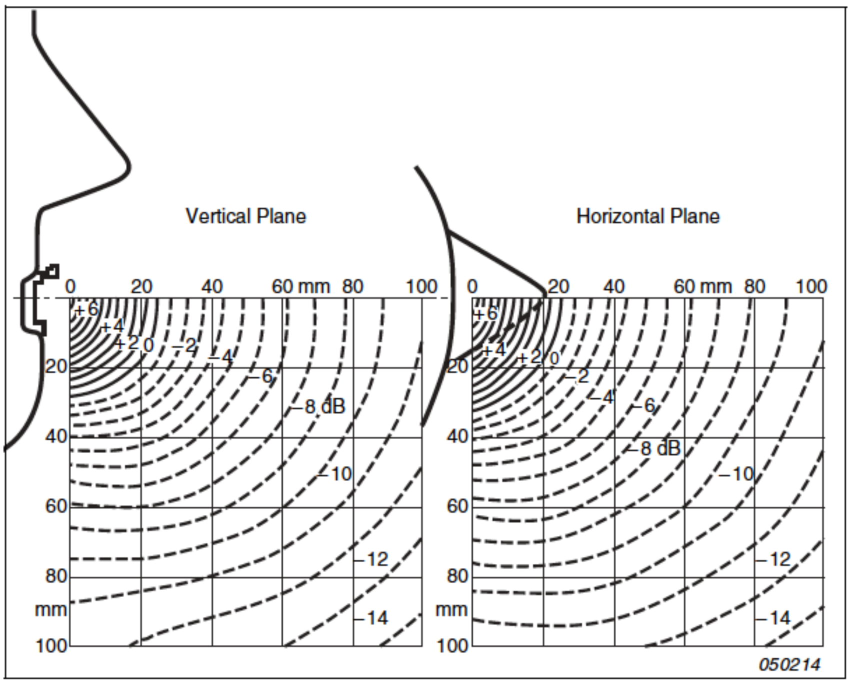

Figure 3.19: The sound pressure level distribution around the mouth opening with the mouth insert in place (average from 300 Hz to 3.3 kHz)[4]

3.4.3 Acquisition Requirements

For the acquisition of the voice signals of the subjects the following equipment has been used:

- a cheek microphone in air (MIPRO MU55-HR) placed at different positions:

1) Aligned to the axis of the mouth, vertically aligned to the tip of the nose, distance 2 cm (Figure 3.20).

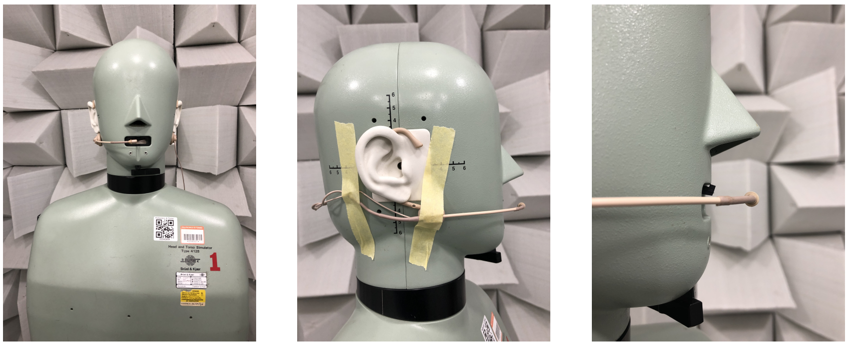

Figure 3.20: Microphone in air aligned to the axis of the HATS mouth.

2) Aligned to the bottom of the nose, 2 cm above the axis of the mouth, vertically aligned to the tip of the nose, distance 2 cm (Figure 3.21).

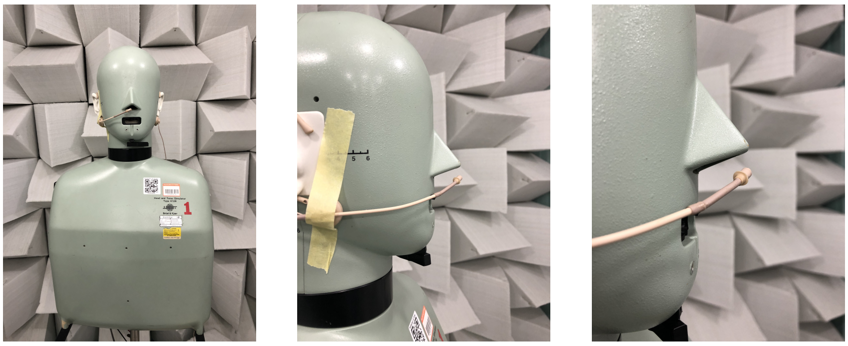

Figure 3.21: Microphone in air aligned to the bottom of the HATS nose.

3) Aligned with the screws on the HATS chin, 2 cm below the axis of the mouth, vertically aligned with the tip of the nose, distance 2 cm (Figure 3.22).

Figure 3.22: Microphone in air aligned with the screws on the HATS chin.

4) Aligned to the corner of the mouth, 4 cm from the axis of the mouth in a lateral position (Figure 3.23).

Figure 3.23: Microphone in air aligned to the corner of the HATS mouth.

- A reference microphone, model M2230 placed at 1 m from the mouth of the HATS (Figure 3.24).

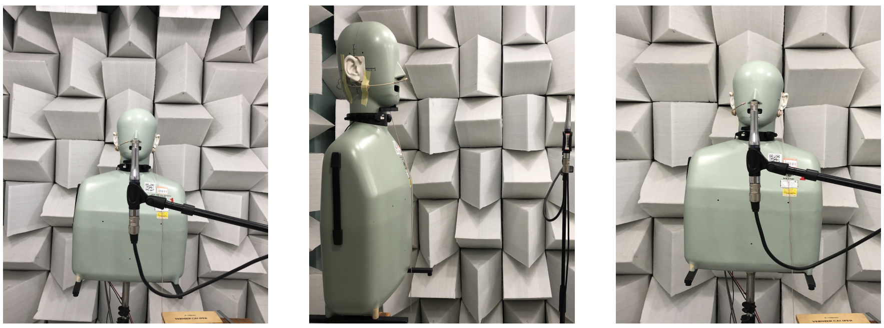

Figure 3.24: Reference microphone at 1 m from the mouth of the HATS.

- An iPhone 8 placed at 40 cm from the mouth of the HATS (Figure 3.25).

Figure 3.25: iPhone 8 at 40 cm from the mouth of the HATS.

The cheek microphone was connected to a portable recorder (EDIROL ROLAND R09-HR), shown in Figure 3.10.
All recordings have been performed with a sampling rate of 44100 Sa/s and a bit-depth of 16 bit.
The lack of the use of the contact microphone is justified by the fact that the "Hats - Head and Torso Simulator" is not equipped with vocal chords.

The HATS, in order to reproduce the artificial vowels, was connected to:

- A car amplifier (Alpine MRP-F200) [23] shown in Figure 3.26, powered by a car battery and characterized by:
1. Selectable 80 Hz high-pass/low-pass crossover at 80 Hz, 18 dB/octave;
2.Preamp outputs and preamp speaker inputs;

Figure 3.26: Alpine MRP-F200 Amplifier

- An audio interface (MOTU Audio Express) [24] shown in Figure 3.27 mainly used to record and mix studio and live performances.

Figure 3.27: MOTU audio express interface

- A Macbook Pro equipped with the Audacity software.

3.4.4 Measuring Chain Evaluation Protocol

The Measuring chain evaluation protocol is structured in three parts:

1. A file has been generated containing three artificial vowels for each subject, in order to obtain only a long file to make easier to record the signals each position of microphone.

2. The long file acquired from the recording devices has been sliced into separate files using an auto-correlation algorithm between the recorded signal and the numeric signal produced by the MC algorithm.

During the measurements an issue with the gain linearity of the evaluation chain has emerged.
This generates offsets and amplitudes which are not compatible.

Figure 3.28, shows a comparison between a vowel obtained artificially and the same vowel after the use of the evaluation chain.

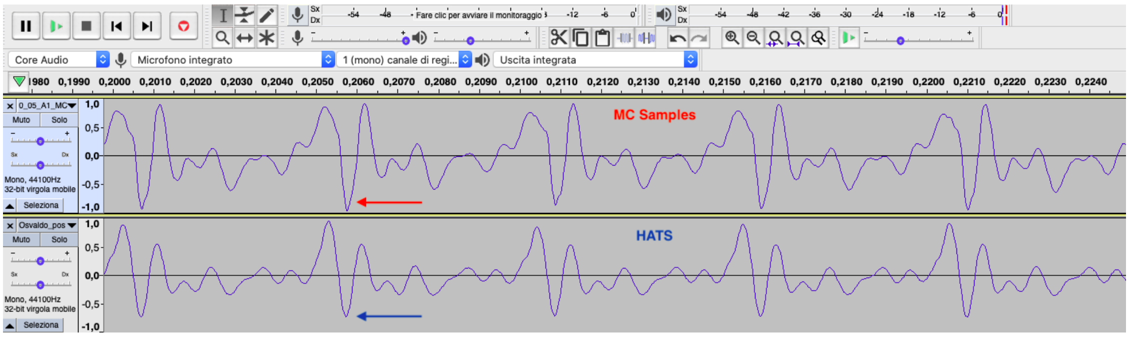

Figure 3.28: Comparison between the MC files and the file obtained in anechoic chamber before the alignment

A difference in the amplitude range of the two signals can be noticed.

In order to attenuate the systematic effect on amplitudes due to the measuring chain non-linearity, an alignment of the amplitudes ranges has been performed. The amplitudes have been normalized accordingly to the artificial signal amplitudes.
No time alignment was necessary because the shift interested only the amplitude.

Figure 3.29 shows a new comparison between the two vowels after the alignment.

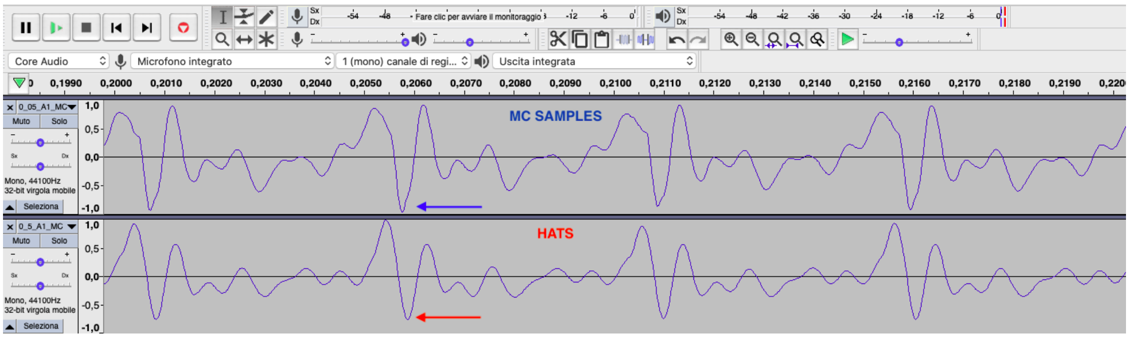

Figure 3.29: Comparison between the MC files and the file obtained in anechoic chamber after the alignment

Even with the alignment done, the offset due to the measuring chain non-linearity is still present.

3. After the amplitude alignment, the same parameters mentioned before have been extracted from the signals.

Chapter 4

Evaluation of the extraction algorithm

4.1 Introduction

One of the goal of the thesis work is to evaluate the reliability of the method used for the extraction algorithm.

4.2 Method description

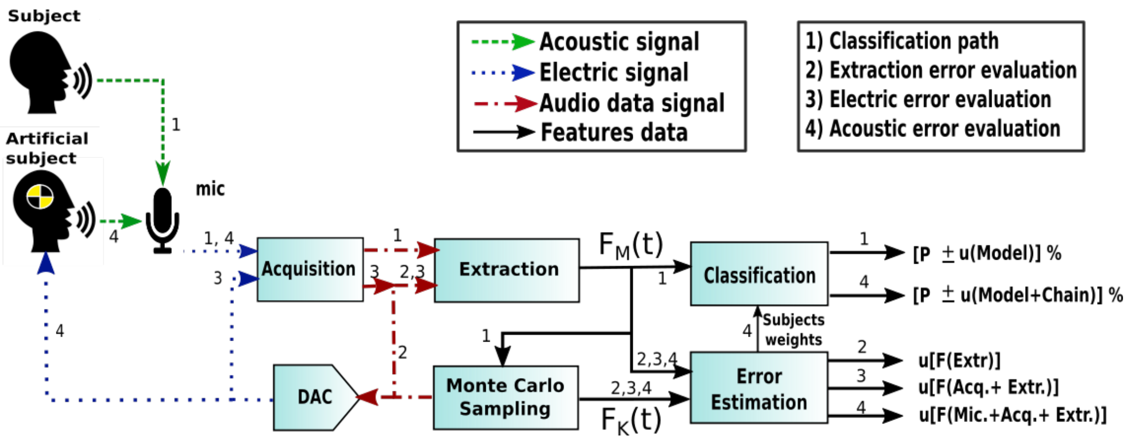

Figure 4.1: Scheme for the error estimation of the extraction algorithm

Figure 4.1 shows the scheme followed to perform the extraction chain evaluation, focusing on steps 1, 2 and 4:

1. The sustained vowel '/a/' performed by a subject has been recorded with a

microphone in air and from these vowels, the sequences of measured periods and amplitudes (T_M, A_M) have been extracted.

2. The measured sequences T_M and A_M have been used to produce known sequences T_K and A_K through the MC algorithm.

4. The artificial vowels (MC Samples), have been reproduced in the anechoic chamber with the HATS and some microphones in different positions and recorded using the same recorder used to acquire the original vowels.

4, 2. From the artificial vowels and the vowels obtained in the anechoic chamber using HATS, the parameters extracted has been used to estimate the relative error considering the known sequences T_K and A_K:

$$Err = \frac{V_m - V_e}{V_e} 100 \tag{4.1}$$

Where V_m is the extracted parameter from the artificial signal, V_e is the parameter calculated starting from the sequences of known periods and amplitudes.
Considering the other extracted parameters, HNR, F_0, RMS and CPPS parameters, the estimated error is:

$$Err = \frac{V_a - V_r}{V_r} 100 \tag{4.2}$$

Where V_a is the extracted parameter from the artificial signal, and V_r is the extracted parameter from the real vocal signal.

The equation 4.2 has been used to calculate parameters such as mean, median, mode, standard deviation, range, 5° percentile, 95° percentile.

For skewness and kurtosis instead, the relative error is not a good measure of uncertainty, due to the fact that they are not linear.

In this case, the absolute error has been calculated as showed in equation 4.3:

$$Err_{abs} = V_a - V_r \tag{4.3}$$

Where V_a is the parameter extracted from the artificial signal, and V_r is the extracted parameter from the real vocal signal.

4.3 Results and method evaluation

The errors of the algorithm of the extraction of parameters, calculated with the expressions 4.1, 4.2 and 4.3, have been used to evaluate the measurement error of the extraction algorithm.
In particular, for each class of patients, a comparison has been done between the average of the errors that each subject has performed.
Afterwards the average errors for every trial have been compared.

4.3.1 Period and amplitude parameters

The Tables in Figures 4.2, 4.3 and 4.4 show relative errors averages obtained from PD, PA and HE subjects and related to time parameters.
The lower values for each parameter are highlighted in light blue.

	Jitt	Jita	RAP	PPQ	vF0
Position 1	14,3	14,3	18,7	13,3	2,7
Position 2	14,1	14,0	18,1	13,3	2,6
Position 3	13,8	13,9	17,9	12,7	2,9
Position 4	14,3	14,0	18,4	13,2	2,4
Telephone	17,3	17,2	21,4	16,6	3,4
Reference Mic	14,5	14,5	18,2	14,1	4,6
Mc_Samples	15,8	15,7	20,4	15,2	2,8

Figure 4.2: Relative Errors Averages for the PD class and related to the parameters to evaluate the frequency stability extracted from the obtained artificial signals. In light blue the lower values for each parameter

	Jitt	Jita	RAP	PPQ	vF0
Position 1	16,5	14,7	19,1	16,5	18,3
Position 2	17,5	16,9	20,1	17,3	11,6
Position 3	17,6	17,1	20,3	17,0	12,3
Position 4	16,1	16,3	18,8	16,2	10,9
Telephone	19,3	19,8	22,0	18,1	13,2
Standard Mic	20,8	19,0	23,7	20,2	20,1
Mc_Samples	17,7	15,7	20,7	18,7	16,0

Figure 4.3: Relative Errors Averages for the PA class and related to the parameters to evaluate the frequency stability extracted from the obtained artificial signals. In light blue the lower values for each parameter

	Jitt	Jita	RAP	PPQ	vF0
Position 1	19,8	19,8	25,6	17,5	2,2
Position 2	19,7	19,7	25,6	17,6	2,3
Position 3	18,6	18,6	24,2	16,6	3,2
Position 4	19,4	19,3	25,1	17,3	6,1
Telephone	25,6	23,9	32,5	22,9	4,1
Reference Mic	23,0	23,0	29,7	20,6	5,3
Mc_Samples	19,5	19,5	25,2	17,5	5,9

Figure 4.4: Relative Errors Averages for the HE class and related to the parameters to evaluate the frequency stability extracted from the obtained artificial signals. In light blue the lower values for each parameter

The Tables in Figures 4.5, 4.6 and 4.7 show the relative errors averages obtained from PD, PA and HE subjects and related to amplitude parameters. The lower values for each parameter are highlighted in light blue.

	Shim	ShdB	APQ	vAm
Position 1	9,74	9,75	8,31	13,01
Position 2	9,51	9,58	8,38	13,48
Position 3	9,97	10,28	8,26	11,94
Position 4	9,81	9,81	8,55	12,53
Telephone	16,65	17,05	11,07	14,32
Reference Mic	8,38	8,68	7,75	17,08
Mc_Samples	2,34	2,69	1,57	0,46

Figure 4.5: Relative Errors Averages for the PD class and related to the parameters to evaluate the amplitude stability of the obtained artificial signals. In light blue the lower values for each parameter.

	Shim	ShdB	APQ	vAm
Position 1	11,4	12,2	9,1	13,4
Position 2	17,1	17,5	14,2	18,3
Position 3	11,1	12,0	9,5	12,7
Position 4	10,9	11,9	9,0	12,6
Telephone	17,6	18,0	13,3	20,4
Standard Mic	10,6	11,3	9,2	14,2
Mc_Samples	4,5	4,8	3,3	1,8

Figure 4.6: Relative Errors Averages for the PA class and related to the parameters to evaluate the amplitude stability of the obtained artificial signals. In light blue the lower values for each parameter.

	Shim	ShdB	APQ	vAm
Position 1	9,7	10,2	5,9	10,3
Position 2	9,7	10,1	5,9	10,9
Position 3	8,9	9,3	5,7	10,4
Position 4	9,4	9,9	5,6	11,7
Telephone	15,5	15,4	10,1	10,7
Reference Mic	13,5	14,3	9,5	12,6
Mc_Samples	1,8	1,9	1,2	1,4

Figure 4.7: Relative Errors Averages for the HE class and related to the parameters to evaluate the amplitude stability of the obtained artificial signals. In light blue the lower values for each parameter.

The errors average of amplitude parameters are lower than those of time parameters.
This happens because there is an epistemic uncertainty in defining a pseudo-period, as they can be defined in different ways, being them, in fact, periodic.
This difficulty in defining the period causes uncertainty in the definition of time parameters unlike amplitude parameters.
When the extraction algorithm finds the periods markers, even a small error in their identification can affects the jitter measurements.
This does not occur for the peak-to-peak amplitudes which remain the same even if the markers shift.
The results obtained are shown in Figure 4.8 for PD subjects, in Figure 4.9 for PA subjects and in Figure 4.10 for HE subjects.

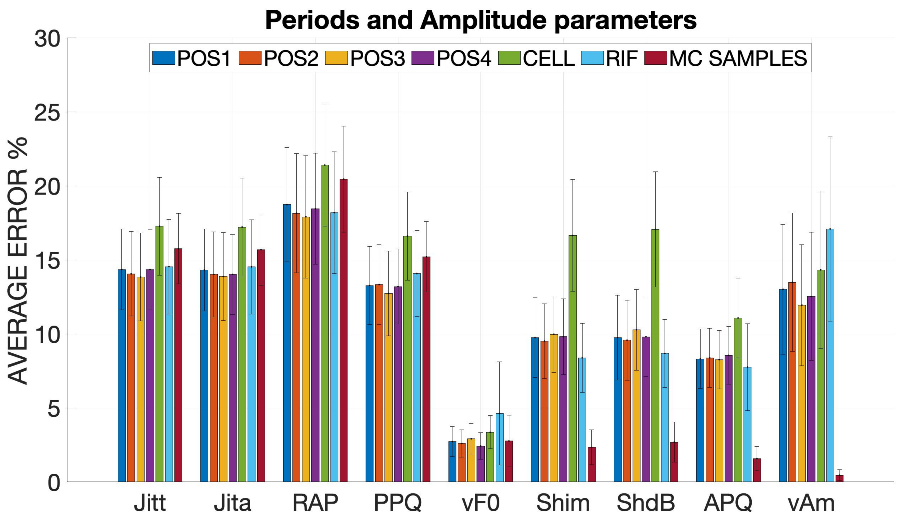

Figure 4.8: Relative Errors Averages % for PD Patients

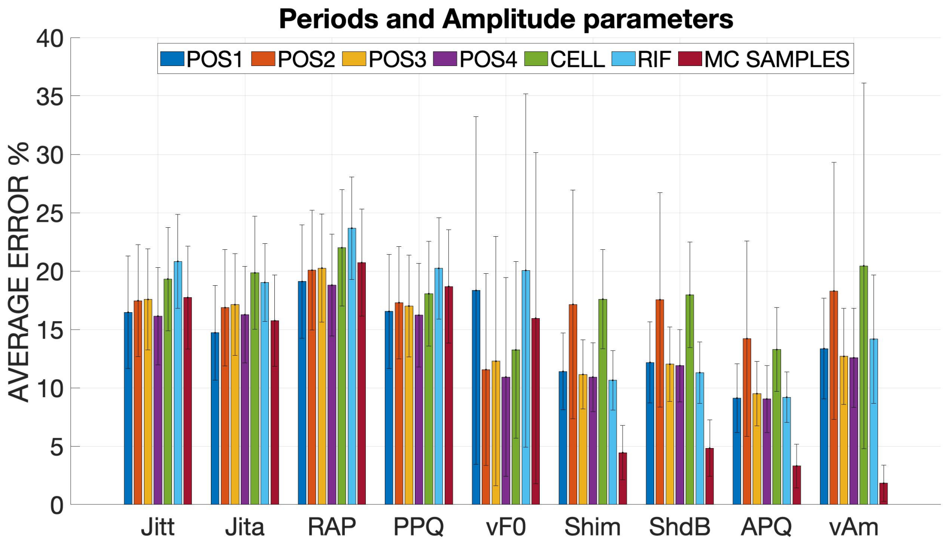

Figure 4.9: Relative Errors Averages % for PA Patients

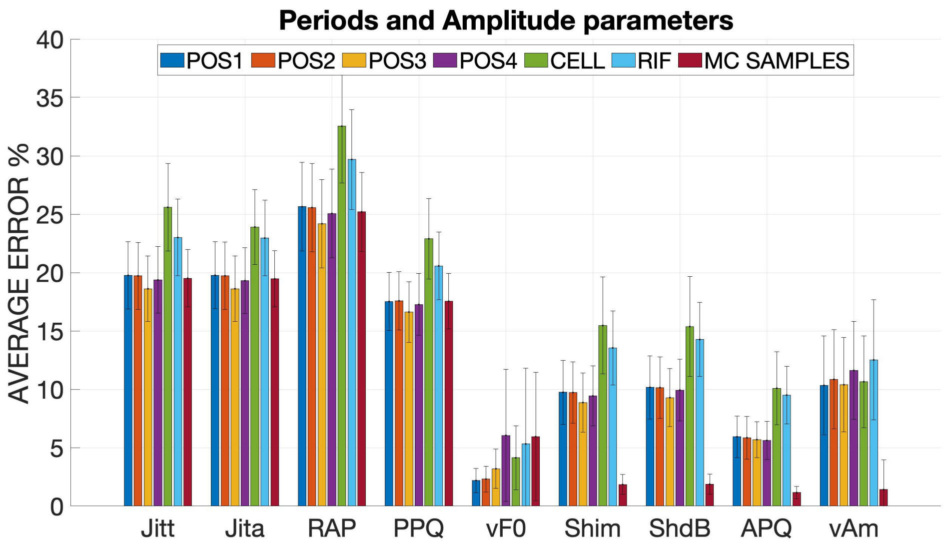

Figure 4.10: Relative Errors Averages % for HE Patients

4.3.2 RMS parameters

The Tables in Figures 4.11, 4.13 and 4.15 show the relative errors averages obtained from PD, PA and HE subjects and related to RMS parameters such as mean, median, mode, standard deviation, range, 5° percentile, 95° percentile.

The Tables in Figures 4.12, 4.14 and 4.16 show the absolute errors obtained from PD, PA and HE subjects and related to RMS parameters such as skewness and kurtosis.
The lower values for each parameter are highlighted in light blue.

RMS	Mean	Median	Mode	Std Dev	Range	5 Percentile	95 Percentile
Position 1	16,2	16,4	18,1	14,5	11,9	17,1	15,8
Position 2	16,8	17,3	19,1	14,9	12,1	17,7	16,3
Position 3	15,6	15,6	17,6	14,0	12,4	16,6	15,3
Position 4	15,8	16,1	17,5	14,8	12,2	16,6	15,6
Telephone	11,8	12,7	13,5	14,3	14,6	12,7	10,9
Reference Mic	17,2	18,0	18,4	16,8	15,5	19,0	16,5
Mc_Samples	2,1	2,6	4,0	2,9	6,8	3,0	2,2

Figure 4.11: RMS Relative Errors Averages for the PD class. In light blue the lower values for each parameter.

RMS	Skewness	Kurtosis
Position 1	0,10	0,33
Position 2	0,11	0,35
Position 3	0,09	0,30
Position 4	0,10	0,34
Telephone	0,19	0,66
Reference Mic	0,18	0,55
Mc_Samples	0,05	0,13

Figure 4.12: RMS Absolute Errors Averages for the PD class, related to skewness and kurtosis parameters. In light blue the lower values for each parameter.

The average of the relative errors of RMS parameters for each class of patients are lower in MC Samples respect to the other recording position.
This behaviour is due to the fact that the error evaluation chain (path 4) has a non linear gain as showed in Section 3.4.4.

When the artificial files were sliced, after being recorded, the offset introduced due to the error evaluation chain (path 4), was still present, even after the removal of the mean and the amplitude alignment.

The results obtained are shown in Figure 4.17 and 4.18 for PD subjects, in Figure 4.19 and 4.20 for PA subjects and in Figure 4.21 and 4.22 for HE subjects.

RMS	Mean	Median	Mode	Std Dev	Range	5 Percentile	95 Percentile
Position 1	12,0	12,5	17,0	11,5	12,6	13,0	12,0
Position 2	12,7	13,9	17,8	16,2	13,0	13,2	12,7
Position 3	12,4	12,6	17,0	11,6	12,3	13,6	12,4
Position 4	12,3	12,8	16,6	12,3	12,4	12,4	12,1
Telephone	19,5	19,6	25,7	21,6	22,6	23,6	19,9
Standard Mic	13,6	13,7	18,0	19,2	15,5	13,4	15,0
Mc_Samples	5,7	5,8	10,5	6,6	8,5	7,5	5,7

Figure 4.13: RMS Relative Errors Averages for the PA class. In light blue the lower values for each parameter.

RMS	Skewness	Kurtosis
Position 1	0,08	0,16
Position 2	0,09	0,16
Position 3	0,08	0,15
Position 4	0,08	0,15
Telephone	0,03	0,04
Standard Mic	0,08	0,14
Mc_Samples	0,07	0,14

Figure 4.14: RMS Absolute Errors Averages for the PA class, related to skewness and kurtosis parameters. In light blue the lower values for each parameter.

RMS	Mean	Median	Mode	Std Dev	Range	5 Percentile	95 Percentile
Position 1	17,0	19,2	20,7	13,1	9,4	24,4	13,3
Position 2	17,0	19,2	20,6	12,8	9,5	24,4	13,3
Position 3	8,2	8,2	10,6	8,7	8,9	8,3	8,0
Position 4	8,9	8,7	10,7	10,9	10,3	8,8	9,5
Telephone	26,0	25,7	29,3	23,0	22,5	26,5	24,8
Reference Mic	12,8	12,3	13,2	17,9	19,4	12,4	13,4
Mc_Samples	2,9	2,7	5,1	4,8	6,4	3,2	3,7

Figure 4.15: RMS Relative Errors Averages for the HE class. In light blue the lower values for each parameter.

RMS	Skewness	Kurtosis
Position 1	0,13	0,29
Position 2	0,13	0,29
Position 3	0,01	0,01
Position 4	0,02	0,05
Telephone	0,04	0,07
Reference Mic	0,02	0,06
Mc_Samples	0,03	0,09

Figure 4.16: RMS Absolute Errors Averages for the HE class, related to skewness and kurtosis parameters. In light blue the lower values for each parameter.

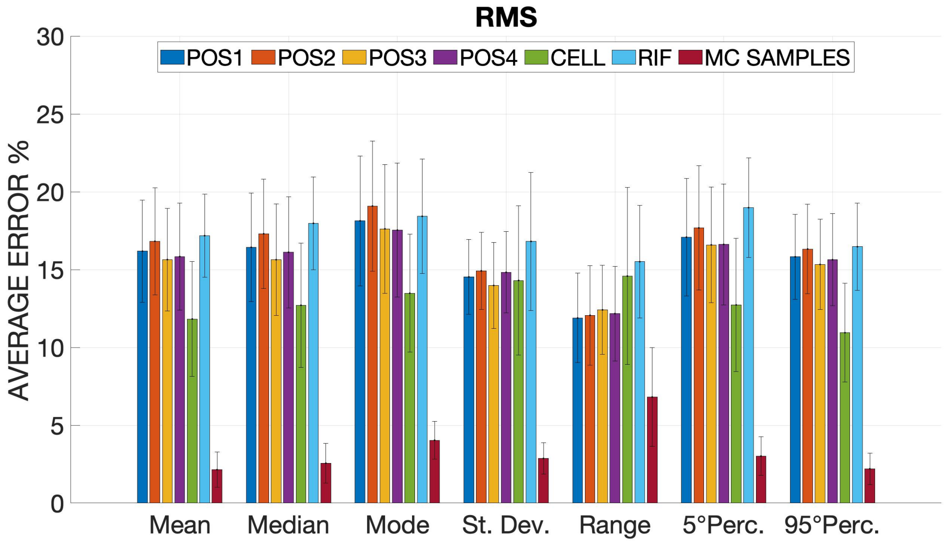

Figure 4.17: RMS Relative Errors Averages % for PD Patients

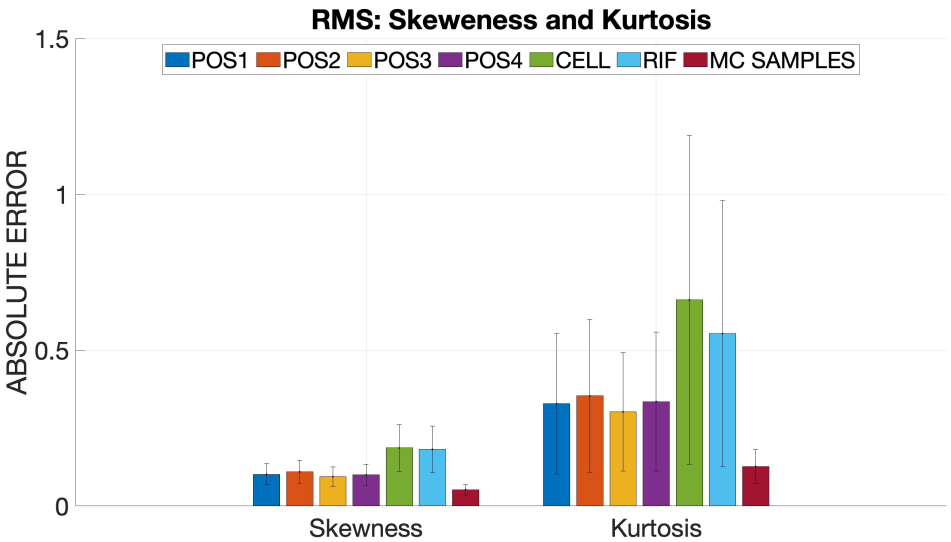

Figure 4.18: RMS Absolute Errors Averages % for PD Patients

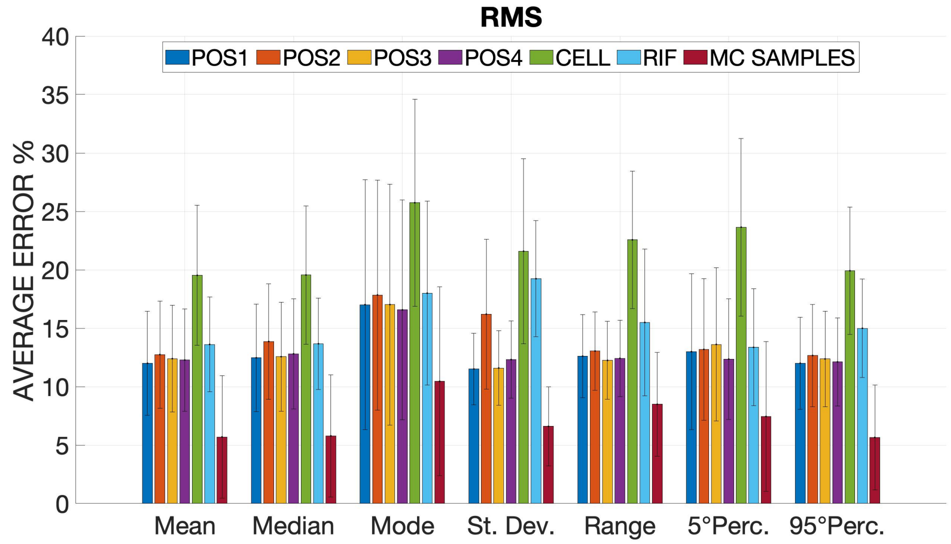

Figure 4.19: RMS Relative Errors Averages % for PA Patients

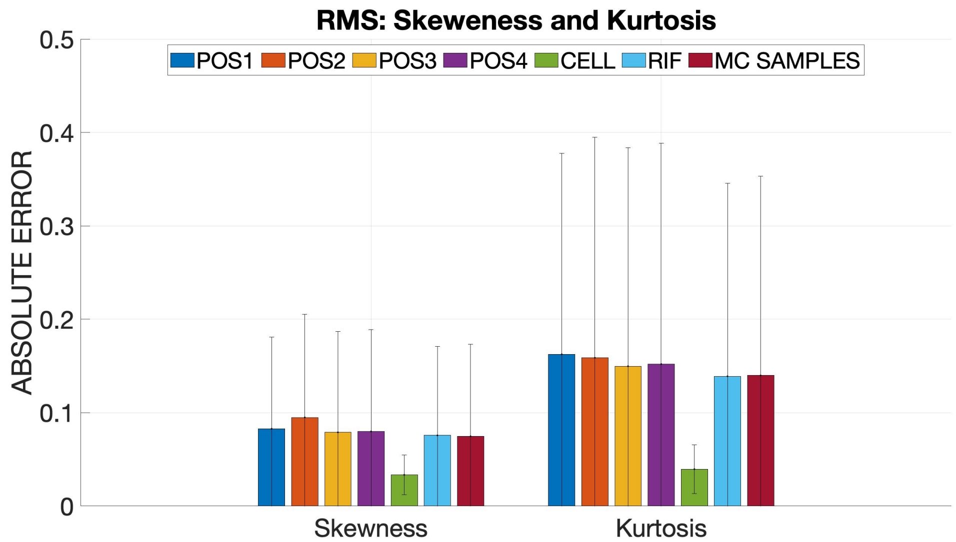

Figure 4.20: RMS Absolute Errors Averages % for PA Patients

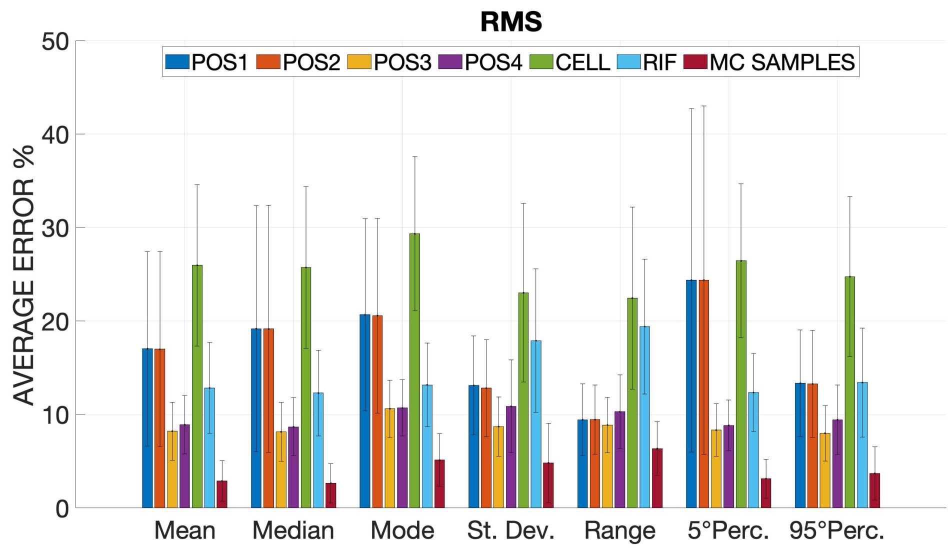

Figure 4.21: RMS Relative Errors Averages % for HE Patients

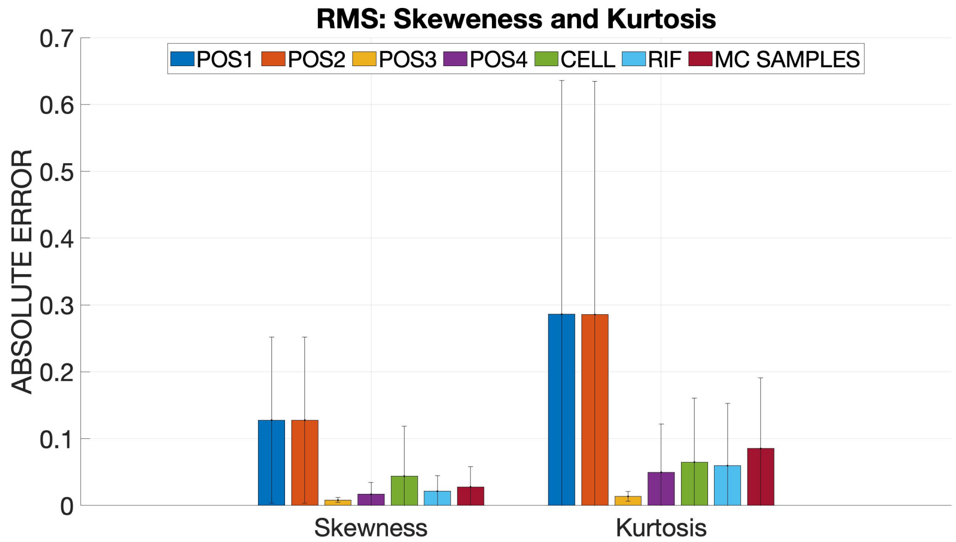

Figure 4.22: RMS Absolute Errors Averages % for HE Patients

4.3.3 CPPS parameters

The Tables in Figures 4.23, 4.25 and 4.27 show the relative errors averages obtained from PD, PA and HE subjects and related to CPPS parameters such as mean, median, mode, standard deviation, range, 5° percentile, 95° percentile.

The Tables in Figures 4.24, 4.26 and 4.28 show the absolute errors obtained from PD, PA and HE subjects and related to CPPS parameters such as skewness and kurtosis.
The lower values for each parameter are highlighted in light blue.

CPPS	Mean	Median	Mode	Std Dev	Range	5 Percentile	95 Percentile
Position 1	4,8	4,9	6,4	12,7	16,6	7,1	3,4
Position 2	4,5	4,6	5,7	13,9	18,4	6,8	3,2
Position 3	4,8	4,7	4,9	12,8	18,3	7,1	3,7
Position 4	4,4	4,4	5,8	14,4	19,6	6,6	3,0
Telephone	9,8	9,7	9,7	41,0	30,5	16,5	5,7
Reference Mic	4,6	4,6	6,2	10,5	18,5	6,3	3,8
Mc_Samples	4,6	4,6	5,8	11,2	12,7	7,1	3,0

Figure 4.23: CPPS Relative Errors Averages for the PD class. In light blue the lower values for each parameter.

CPPS	Skewness	Kurtosis
Position 1	0,24	0,83
Position 2	0,25	0,98
Position 3	0,27	0,92
Position 4	0,23	0,85
Telephone	0,35	1,11
Reference Mic	0,40	0,99
Mc_Samples	0,27	1,15

Figure 4.24: CPPS Absolute Errors Averages for the PD class, related to skewness and kurtosis parameters. In light blue the lower values for each parameter.

CPPS	Mean	Median	Mode	Std Dev	Range	5 Percentile	95 Percentile
Position 1	5,1	4,6	11,8	13,6	15,3	6,4	5,9
Position 2	6,6	6,3	12,8	14,1	15,7	8,4	6,7
Position 3	5,2	4,9	10,8	13,2	14,5	6,2	6,0
Position 4	5,2	4,7	11,4	13,1	16,1	6,5	6,2
Telephone	10,4	11,0	16,1	22,1	20,5	12,6	7,9
Standard Mic	7,8	7,5	11,7	14,8	15,9	14,5	6,8
Mc_Samples	6,0	6,0	11,1	13,8	14,9	8,8	5,3

Figure 4.25: CPPS Relative Errors Averages for the PA class. In light blue the lower values for each parameter.

CPPS	Skewness	Kurtosis
Position 1	0,05	0,07
Position 2	0,04	0,08
Position 3	0,03	0,07
Position 4	0,03	0,07
Telephone	0,03	0,05
Standard Mic	0,03	0,07
Mc_Samples	0,03	0,07

Figure 4.26: CPPS Absolute Errors Averages for the PA class, related to skewness and kurtosis parameters. In light blue the lower values for each parameter.

CPPS	Mean	Median	Mode	Std Dev	Range	5 Percentile	95 Percentile
Position 1	5,3	5,4	7,2	12,8	15,1	7,8	3,7
Position 2	5,5	5,6	7,4	13,2	14,9	8,1	4,1
Position 3	3,4	3,6	4,6	9,8	11,7	4,6	2,7
Position 4	4,6	4,7	4,6	10,5	15,7	5,3	3,9
Telephone	2,7	3,0	4,9	33,3	24,4	4,6	4,4
Reference Mic	4,5	4,6	4,7	8,7	16,5	5,8	3,6
Mc_Samples	2,8	2,8	4,6	7,0	12,8	3,8	2,2

Figure 4.27: CPPS Relative Errors Averages for the HE class. In light blue the lower values for each parameter.

CPPS	Skewness	Kurtosis
Position 1	0,039	0,042
Position 2	0,038	0,042
Position 3	0,003	0,008
Position 4	0,005	0,008
Telephone	0,011	0,018
Reference Mic	0,006	0,010
Mc_Samples	0,006	0,009

Figure 4.28: CPPS Absolute Errors Averages for the HE class, related to skewness and kurtosis parameters. In light blue the lower values for each parameter.

The explanation on the distribution of the values obtained can be found considering the fact that unlike for parameters such as jitter and shimmer, calculated with respect to a reference value, a reference value for CPPS does not exist.

CPPS is a dispersion measurement, extracted from the original signal.
For this reason an absolute estimation of the measurement error between the original and the artificial signals can't be done.

The worst relative errors have been obtained with the iPhone 8.
This is probably due to the file compression of the audio recording app (Memo vocali) which alter the spectral quality of the recorded material.

The results obtained are shown in Figure 4.29 and 4.30 for PD subjects, in Figure 4.31 and 4.32 for PA subjects and in Figure 4.33 and 4.34 for HE subjects.

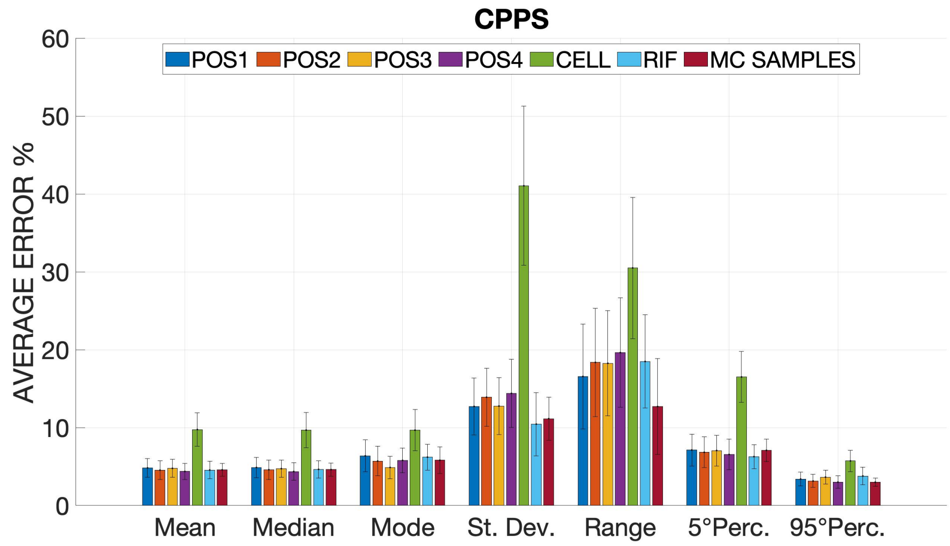

Figure 4.29: CPPS Relative Errors Averages % for PD Patients

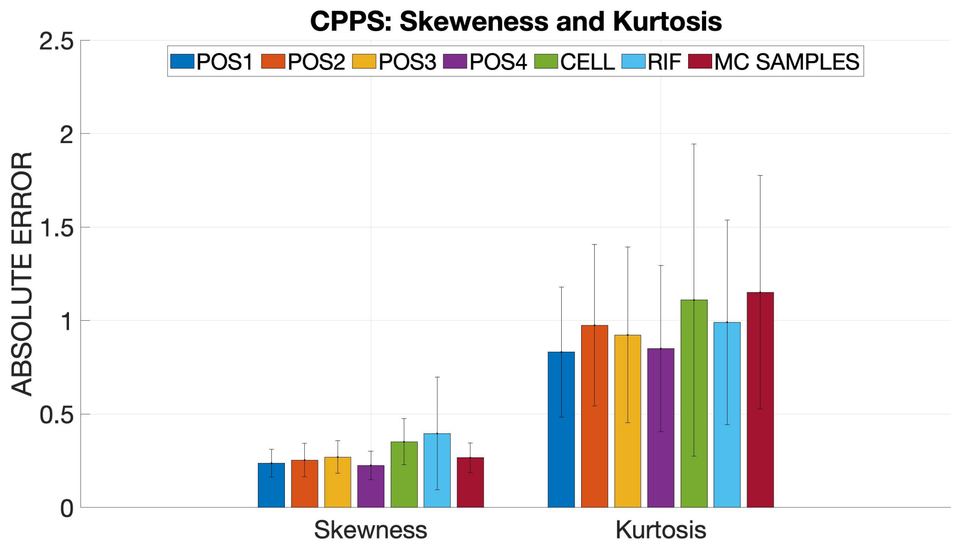

Figure 4.30: CPPS Absolute Errors Averages % for PD Patients

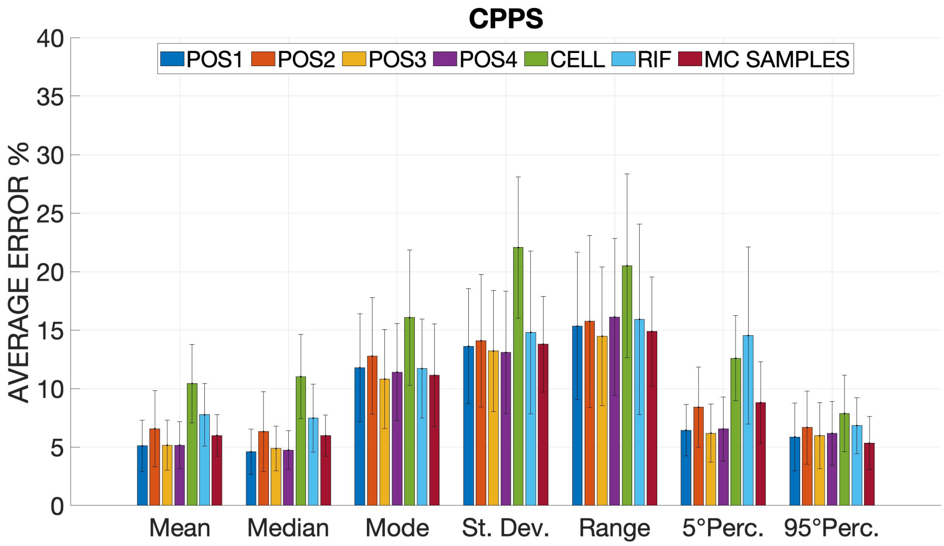

Figure 4.31: CPPS Relative Errors Averages % for PA Patients

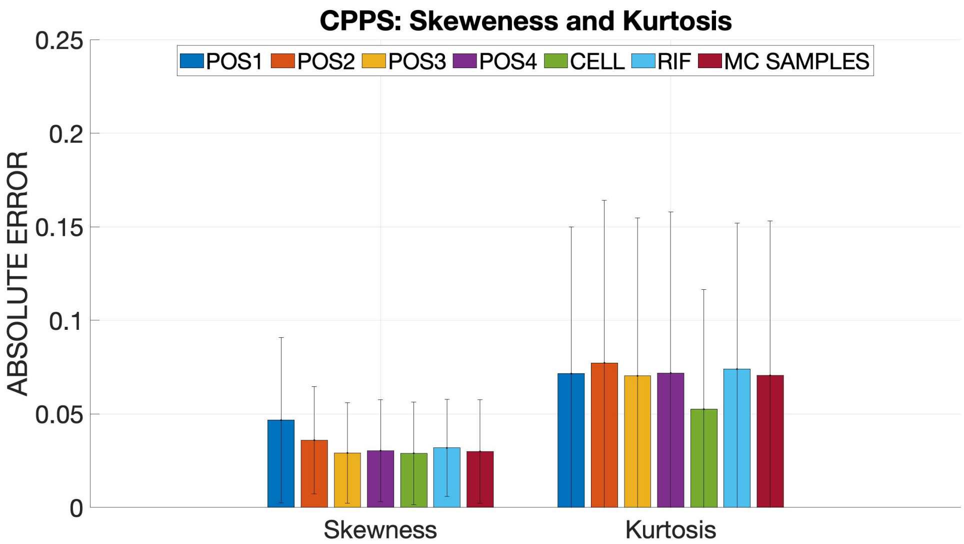

Figure 4.32: CPPS Absolute Errors Averages % for PA Patients

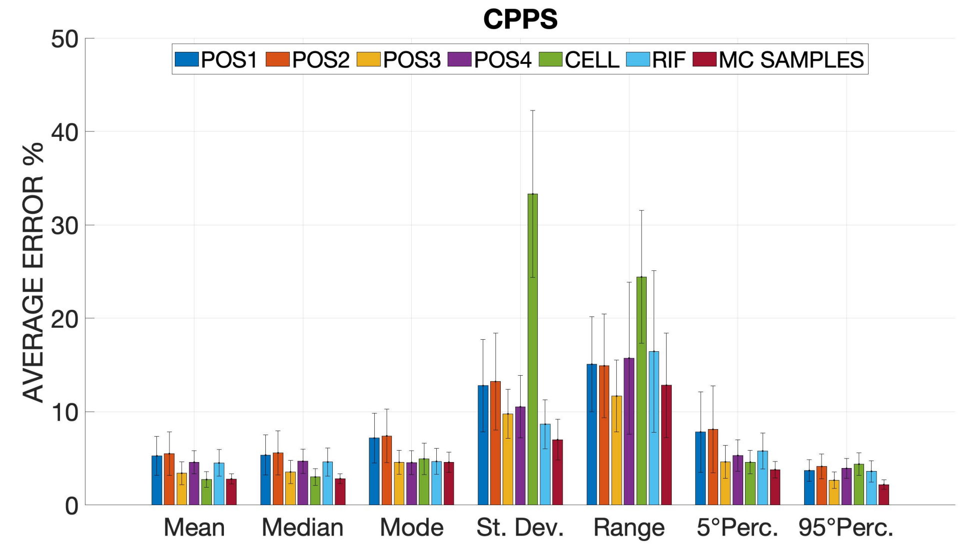

Figure 4.33: CPPS Relative Errors Averages % for HE Patients

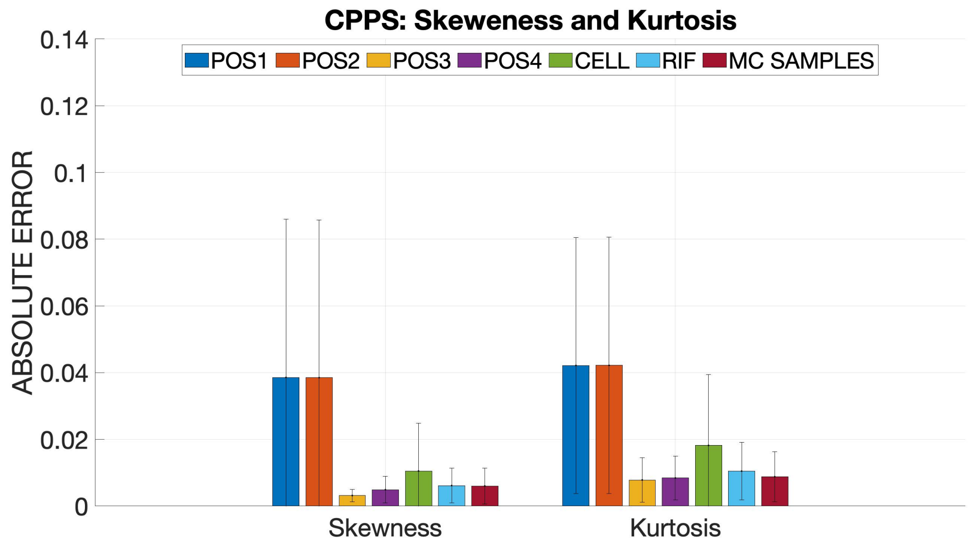

Figure 4.34: CPPS Absolute Errors Averages % for HE Patients

Chapter 5

Classification methods

5.1 Introduction

To evaluate the reliability of the features that constitute the dataset a feature selection algorithm has been used.
Furthermore, to discriminate the Parkinson's disease patients from healthy subjects and pathological non-parkinsonian patients, a weighted classification algorithm has been developed using the measurement errors evaluated in the previous sections.

5.2 Feature selection

A feature selection algorithm, allows to reduce the size of the initial dataset, providing a subset consisting of features only relevant and uncorrelated [25].
To remove related features, has been used the Correlation-Based Feature Selection (CBFS) method.

This algorithm excludes the correlation between variables as a measure of the reliability of the features.

The correlation index assumes values between -1 and 1, where a value of 1 indicates variables totally correlated, a value of 0 indicates totally uncorrelated variables and a value equal to -1 indicates totally inversely correlated variables.

If the correlation is higher than a certain threshold it means that the combination will not have a high predictive power [26].

In order to evaluate the reliability of the feature, considering as useful information, their non-correlation, the squared correlation coefficient has been used.

Considering HE, PA and PD classes, the correlation matrix r_{ij} and the corresponding p-value p_{ij} have been calculated.

The Pearson correlation index, defined as the covariance of two variables X and Y divided by the product of the standard deviations of the two variables [27]:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{5.1}$$

Where σ_{XY} is the covariance between X and Y, σ_X is the standard deviation of X and σ_Y is the standard deviation of Y. The matrices r_{ij} and p_{ij} were used to calculate the candidate matrix:

$$candidates_{ij} = [r_{ij}^2 LM_{ij} + \overline{LM}_{ij}] \tag{5.2}$$

Where LM is a binary matrix where the ij cell depends on the p-value of the ij correlation.
$\overline{LM}_{ij}$ is the logical complement of LM_{ij}.
$LM_{ij} = 1$ if p_{ij}<0.05 and $LM_{ij} = 0$ otherwise.
The $candidates_{ij}$ value is equal to 1 if the p-value is greater than 0.05, which it means that the correlation (or non-correlation) is not significant.

The lowest values of the candidate matrix are relative to uncorrelated and therefore non-redundant features.

For this work the threshold value has been chosen equal to 0.6.

5.3 The logistic regression

The logistic regression is a statistical model that examines the relationship between independent variables and a dichotomous dependent variable, that assumes, as values, only 0 or 1, providing a probability via a logistic function [28]. This function is represented as an "S" shaped curve and is defined by the equation:

$$f(x) = \frac{1}{1 + e^{f(x)}} \tag{5.3}$$

where $f(x)$ is a generic linear combination of features and coefficients. An example is shown in Figure 5.1

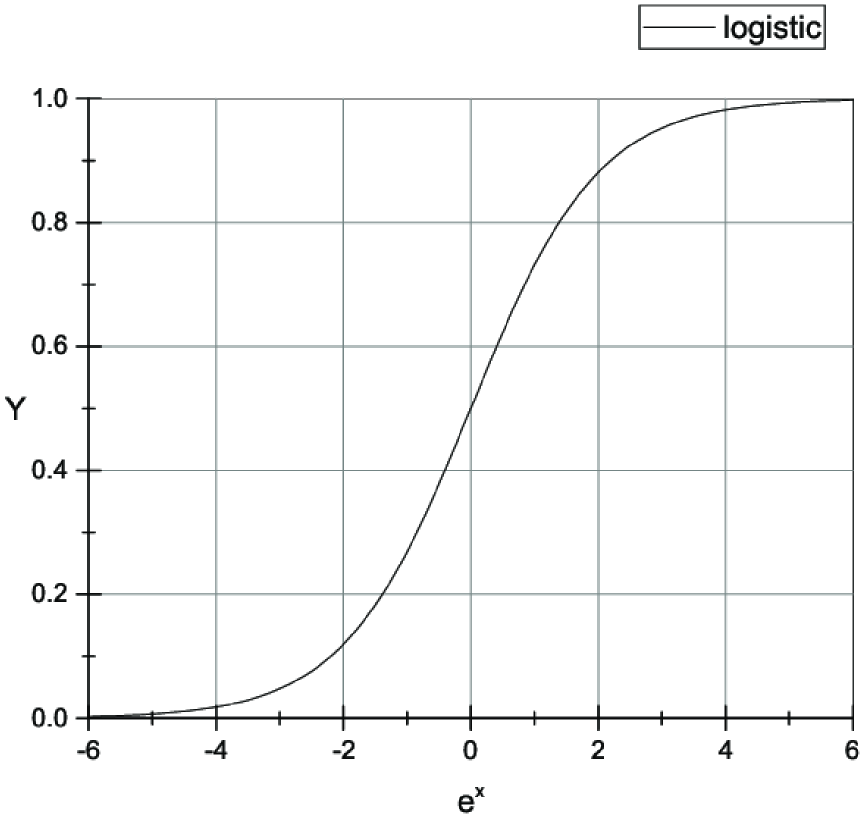

Figure 5.1: Example of a logistic function for $f(x) = x$. [5]

This model used, provides the probability of belonging to the healthy class according to the expression:

$$P_{HE} = \frac{1}{1 + e^{-(\beta_0 + \sum_i \beta_i x_i)}} \tag{5.4}$$

Where P_{HE} is the probability to belong to healthy class of patients and β_i are the coefficients of linear combination with features x_i, considering, as threshold for classification $P_{HE} = 0.5$.

To evaluate the performance of the classifier, the confusion matrix has been used. In this table, the columns represent the true class and the rows represent the class predicted by the classifier.

An example is shown in Figure 5.2

	True Classification	
Classifier Result	P	N
P	TP	FP
N	FN	TN

Figure 5.2: Confusion Matrix.

The meaning of the acronyms in the table is:

- True Positive (TP): number of elements correctly classified as positive (P);

- True Negative (TN): number of elements correctly classified as negative (N);

- False Positive (FP): number of elements classified as positive but that are actually negative;

- False Negative (FN): number of elements classified as negative but that are actually positive.

The consideration that can be done, regarding the table, is that the cells that are on the diagonal of this table indicate the number of elements that have been correctly classified considering their belonging class.

A parameter used to determine the performance of the considered model is the Standard Accuracy.
It is calculated as the number of items that have been correctly classified divided by the total number of classified:

$$Acc = \frac{(TP + PN)}{(TP + TN + FP + FN)} 100 \tag{5.5}$$

The standard accuracy range goes from 0% which is the worst accuracy obtainable to 100% which is the best accuracy obtainable.

5.4 The Combinatorial algorithm

The combinatorial algorithm selects a group of features to train the classification model.
For this work, groups of 2 to 4 features have been used:

1. Determine the number of combinations k of n features in p combinations with the following formula:

$$k = \frac{n!}{p!(n-p)!} \tag{5.6}$$

2. For each not repeated combination of p features do:

2.1. Check the ij cell of the candidates matrix:
- if $candidates_{ij} > 0.6$, then the combination is not considered valid;
- if $candidates_{ij} <= 0.6$ go to step 3:

3. Train of a logistic regression model:
- if the p-value of model coefficients <0.05, the model is valid;
- else:
- go to step 2.;

5.5 Weighted Logistic Regression

For this work a weighted logistic regression algorithm is proposed to discriminate voices of PD patients from the voices of HE patients and the voices of PD patients from the voices of PA patients.

The weights have been defined according to the errors evaluated in the previous sections.

The combination of features considered by the classifier have been selected in order to give more weight to subjects whose features have been extracted with less errors.

The relation between errors and weight has been choosen as the reciprocal of the relative error:

the value of the weight w was taken equal to the reciprocal of the error err:

$$w = \frac{1}{err} \tag{5.7}$$

To estimate the uncertainty associated with the equation 5.4, the following equation can be used:

$$u(p) = \sqrt{\sum_{i=1}^{N} (\frac{\delta p}{\delta x_i})^2 u^2(x_i)} \tag{5.8}$$

Where p is the predicted probability, x_i is the i-th feature and $u(x_i)$ is the i-th feature uncertainty.

To calculate the uncertainty of P_{HE}, applying the equation 5.4, the expression is:

$$u(P) = \frac{e^{\beta_0 + \sum_{i=1}^{N} \beta_i f_i}}{(1 + e^{\beta_0 + \sum\limits_{i=1}^{N} \beta_i f_i})^2} \sqrt{u^2(\beta_0) + \sum_{i=1}^{N} f^2 u^2(\beta_i) + \beta_i u^2(f_i)} \tag{5.9}$$

Where $u(\beta_i)$ is the uncertainty of the coefficients and $u(f_i)$ is the uncertainty of the features.

To estimate the performance of the classifier, a new metric has been proposed.
Such metric is based on the classification rate of the subjects which excludes the non-classified subjects from the accuracy estimation.
The non-classified element has a probability range obtained from the confidence interval that crosses the classification threshold $P_{HE} = 0.5$.
An example is shown in Figure 5.3 and Figure 5.4.
Using the "sortrows" matlab function, the probabilities have been sorted in ascending order.

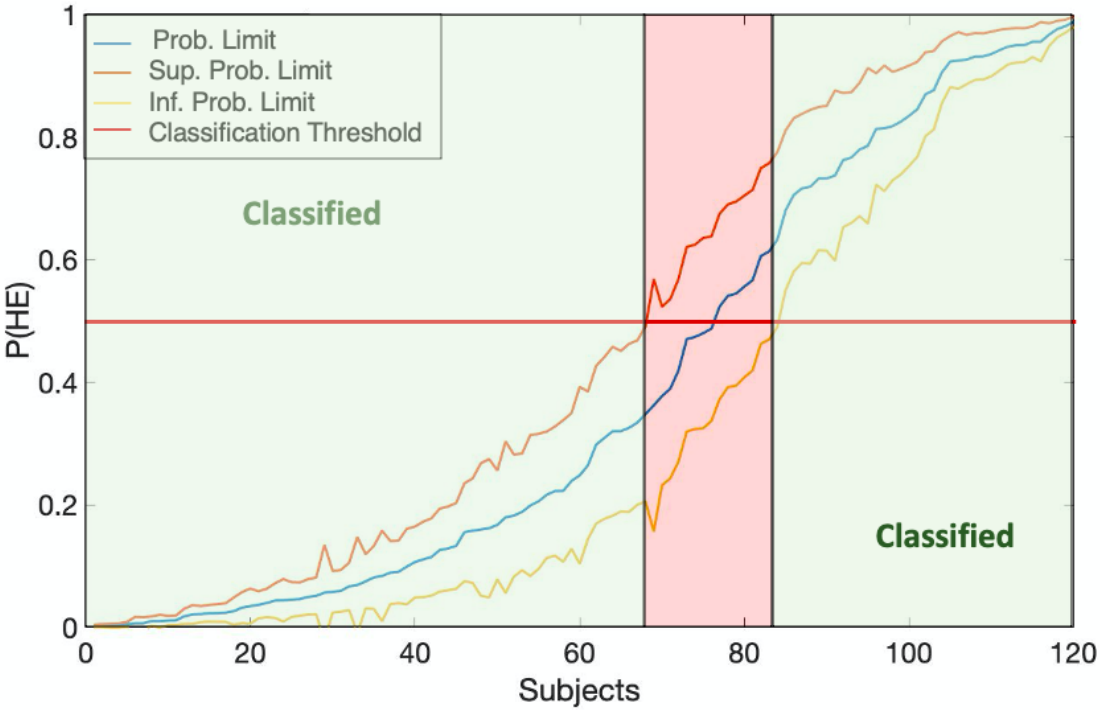

Figure 5.3: Probability of belonging to HE class with relative confidence limits. In the green area the classified elements, in the red one the non-classified elements.

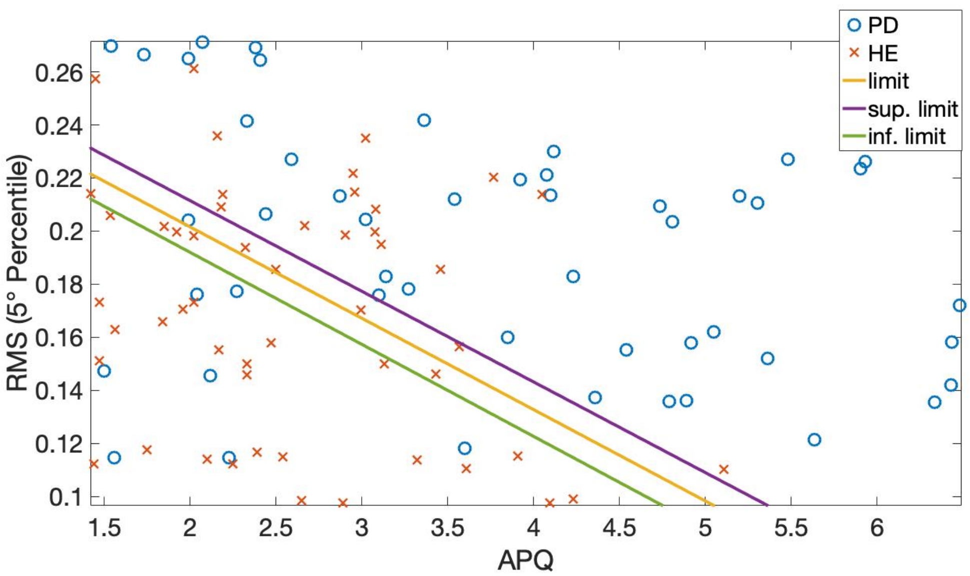

Figure 5.4: Probability of belonging to HE class with relative confidence limits

Three parameters have been defined:
- Fraction of Classified;
- Pessimistic Accuracy;
- Optimistic Accuracy;

The Fraction of Classified is calculated as:

$$Class = \frac{N_C}{N_{TOT}} 100 \tag{5.10}$$

Where N_C is the number of classified element and N_{TOT} is the total number of elements.

The Pessimistic Accuracy is calculated as the sum of true positives and true negatives of the classified elements N_{TC} divided by the total number of elements N_{TOT}:

$$Acc_{(Class)p} = \frac{N_{TC}}{N_{TOT}} 100 \tag{5.11}$$

The Optimistic Accuracy is calculated as the sum of true positives and true negatives of the classified elements N_{TC} divided by the total number of classified elements $N_{ClassTOT}$:

$$N_{ClassTOT} = \frac{Class * N_{TOT}}{100} \tag{5.12}$$

$$Acc_{(Class)o} = \frac{N_{TC}}{N_{ClassTOT}} 100 \tag{5.13}$$

Chapter 6

Results and discussion

6.1 Introduction

Using a weighted logistic regression model, the main goal of the classification is to:

- Discriminate PD from HE subjects;

- Discriminate PD from PA subjects;

The features extracted from the real signal and the artificial signal, together with the measurement error, have been used to obtain two different subsets:

1. Classification 1:
• the elements to be classified are the average of the 6 vowels (3 real + 3 artificial) associated to each subject and the errors are the standard deviation of the average of the 6 vowels.
This classification was repeated seven times, considering as artificial vowels, the vowels obtained directly from the MonteCarlo method (MC Samples) and those recorded in the anechoic chamber.

2. Classification 2:
• the elements to be classified are the average of the 15 vowels (3 real + 12 artificial considering those obtained from the recordings in the anechoic chamber considering 4 positions of the microphone in air) associated with each subject and the errors are the standard deviation of the average of the 15 vowels.

The Table in Figure 6.1 shows the two classification methods.

CLASSIFICATION	ELEMENT CONSIDERED	ERROR
Classification 1	$\overline{3VOWELS_{R+A}}$	$\frac{\sigma_{R+A}}{\sqrt{6}}$
Classification 2	$\overline{3VOWELS_{R}+12VOWELS_{A}}$	$\frac{\sigma_{R+A}}{\sqrt{15}}$

Figure 6.1: Classification Methods.

6.2 PD vs HE subjects

6.2.1 Classification 1

The classification 1 has been done considering the average of 3 real vowels and 3 artifical vowels, using as errors, the standard deviation of the average of the 6 vowels.
In this case, there is a reduction of the dataset from 30 to 10 elements for each class of subjects.

The Table in Figure 6.2 shows the best performances obtained considering the first classification method, for each microphone position and a number of feature equal to 2.

CLASSIFICATION 1: PD VS HE - 2 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1	9 30	89.47	68.42	68.42	100
POSITION 2	9 30	94.74	68.42	68.42	100
POSITION 3	9 30	94.74	73.68	73.68	100
POSITION 4	9 30	94.74	73.68	73.68	100
IPHONE 8	9 30	94.74	73.68	73.68	100
REF. MIC.	9 30	89.47	68.42	68.42	100
MC_SAMPLES	9 33	85	45	45	100

Figure 6.2: Performances of classification 1, discriminating PD vs HE patients, for a number of features equal to 2.

Considering classification 1 for two features, the best performances have been obtained in Position 2, 3, 4 and with the iPhone 8 as regards the Standard Accuracy while in Position 3, 4 and with the iPhone 8 as regards the Fraction of Classified and the Pessimistic Accuracy.
The Optimistic Accuracy, has reached values up to 100% in every position.
The features in almost every position are the same, showing that there was a good reproducibility of measurements even if only one feature differs from the others in MC Samples.

The Table in Figure 6.3 shows the best performances obtained for the first classification method, for each microphone position and a number of feature equal to 3.

CLASSIFICATION 1: PD VS HE - 3 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1	1 39 40	89.47	36.84	31.58	85.71
POSITION 2	1 39 40	89.47	36.84	31.58	85.71
POSITION 3	1 40 41	78.95	42.11	42.11	100
POSITION 4	1 39 40	89.47	42.11	36.84	87.50
IPHONE 8	6 9 33	78.95	42.11	42.11	100
REF. MIC.	1 7 31	84.21	42.11	42.11	100
MC_SAMPLES	1 40 41	78.95	42.11	42.11	100

Figure 6.3: Performances of classification 1, discriminating PD vs HE for each microphone position and a number of feature equal to 3.

Considering classification 1 for three features, the best performances have been obtained in the Position 1, 2 and 4 as regards the Standard Accuracy while the worst performances have been obtained in Position 3, with the iPhone 8 and MC Samples.

Considering the Fraction of Classified, the best performances have been obtained in Position 3, 4, with the iPhone 8, Reference Microphone and MC Samples.

Considering the Pessimistic Accuracy, the best performances have been obtained in Position 3, with the iPhone 8, Reference Microphone and MC Samples. The same results have been obtained considering the Optimistic Accuracy, that in those positions has reached values up to 100%.

In this case, the features considered are more dishomogeneous than those obtained

in the previous classification.
While the percentage of the Optimistic Accuracy hasn't changed, with three features there is a reduction of the other parameters accuracies.

The Table in Figure 6.4 shows the best performances obtained for the first classification method, for each microphone position and a number of feature equal to 4.

CLASSIFICATION 1: PD VS HE - 4 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1	4 7 15 16	94.74	68.42	68.42	100
POSITION 2	4 7 15 16	94.74	68.42	68.42	100
POSITION 3	1 12 40 41	94.74	73.68	73.68	100
POSITION 4	1 12 40 41	89.47	73.68	73.68	100
IPHONE 8					
REF. MIC.	1 12 40 41	84.21	31.58	26.32	83.33
MC_SAMPLES	1 12 15 40	85	50	45	90

Figure 6.4: Performances of classification 1, discriminating PD vs HE for each microphone position and a number of feature equal to 4.

Considering classification 1 for four features, the best performances have been obtained in Position 1, 2 and 3 as regards the Standard Accuracy.

Considering the Fraction of Classified and the Pessimistic Accuracy, the best performances have been obtained in Position 3 and 4 while considering the Optimistic Accuracy, the best performances have been obtained in Positions 1, 2, 3 and 4.

The worst performances have been obtained with the Reference Microphone.
In this case, the features considered are more dishomogeneous than those obtained in the previous classification.

Particular attention have to be payed because the feature selection algorithm failed to find a valid feature combination using the chosen threshold values for the iPhone 8. This is due to the fact that the increasing number of features leads to a rise in classification error.
This happens because this is an additive uncertainty model in which the uncertainty contributions are added quadratically.

6.2.2 Classification 2

The classification 2 was done considering the average of the 15 vowels (3 real + 12 artificial considering those obtained from the recording in the anechoic chamber in 4 positions of the microphone in air) associated with each subject and the errors are the standard deviation of the average of the 15 vowels.
Even in this case there is a reduction of the dataset considered from 30 to 10 elements.

The Table in Figure 6.5 shows the best performances obtained considering the second classification method for a number of feature equal to 2.

CLASSIFICATION 2: PD VS HE - 2 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1:4	9 33	94.74	73.68	73.68	100

Figure 6.5: Performances of classification 2, discriminating PD vs HE for a number of feature equal to 2.

Considering classification 2 for two features, the percentages of the Standard Accuracy, Fraction of Classified, Pessimistic and Optimistic Accuracy are equal to those of the classification 1.
In this classification, the same features obtained considering MC Samples in classification 1, have been selected.

The Table in Figure 6.6 shows the best performances obtained considering the second classification method for a number of feature equal to 3.

CLASSIFICATION 2: PD VS HE - 3 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1:4	27 32 36	89.47	68.42	63.16	92.31

Figure 6.6: Performances of classification 2, discriminating PD vs HE for a number of feature equal to 3.

The values of the parameters for three features are lower with respect to those obtained for two features.
The features considered before, have not been considered in this case, even the Optimistic Accuracy hasn't reached values up to 100%.

The Table in Figure 6.7 shows the best performances obtained for the second classification method for a number of feature equal to 4.

CLASSIFICATION 2: PD VS HE - 4 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1:4	1 2 6 15	84.21	63.16	63.16	100

Figure 6.7: Performances of classification 2, discriminating PD vs HE for a number of feature equal to 4.

In this case, the percentages of the Standard Accuracy, Fraction of Classified and Pessimistic Accuracy are lower than classification 2 for three features while the percentage of Optimistic Accuracy has increased and reached values up to 100%.

6.3 PD vs PA subjects

6.3.1 Classification 1

The Table in Figure 6.8 shows the best performances obtained considering the first classification method, for each microphone position and a number of feature equal to 2.

CLASSIFICATION 1: PD VS PA - 2 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1	23 37	95	60	55	91.67
POSITION 2	23 37	95	60	55	91.67
POSITION 3	23 37	95	60	55	91.67
POSITION 4	23 37	95	60	55	91.67
IPHONE 8	23 37	95	60	55	91.67
REF. MIC.	23 37	95	60	55	91.67
MC_SAMPLES	23 37	95	60	55	91.67

Figure 6.8: Performances of classification 1, discriminating PD vs PA for each microphone position and a number of feature equal to 2.

The classification 1 for two features, highlights the same performances obtained considering different positions of microphones.
In this case the accuracies show that there is a good reproducibility of the measurements.

The Table in Figure 6.9 shows the best performances obtained for the first classification method, for each microphone position and a number of feature equal to 3.

CLASSIFICATION 1: PD VS PA - 3 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1	23 41 42	90	50	50	100
POSITION 2	23 41 42	90	50	50	100
POSITION 3	23 41 42	90	50	50	100
POSITION 4	23 37 40	95	60	60	100
IPHONE 8	6 23 38	95	50	50	100
REF. MIC.	23 41 42	90	50	50	100
MC_SAMPLES	7 23 38	95	50	50	100

Figure 6.9: Performances of classification 1, discriminating PD vs PA for each microphone position and a number of feature equal to 3.

Considering classification 1 for three features, the best performances have been obtained in Position 4 regarding the Fraction of Classified and the Pessimistic Accuracy.

Considering the Standard Accuracy, the best performances have been obtained in Position 4, with the iPhone 8 and MC Samples, while the Optimistic Accuracy has reached values up to 100% in every position.

The features highlighted are the same for Positions 1, 2, 3 of the microphone in air and for the Reference Microphone while in the other positions there are other features considered.

Table in Figure 6.10 shows the best performances obtained considering the first classification method, for each microphone position and a number of feature equal to 4.

CLASSIFICATION 1: PD VS PA - 4 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1	8 22 39 40	75	40	40	100
POSITION 2	7 23 39 46	95	50	50	100
POSITION 3	6 23 39 46	90	45	45	100
POSITION 4	15 23 41 42	90	50	50	100
IPHONE 8	7 23 39 46	95	45	45	100
REF. MIC.	15 23 41 42	90	50	50	100
MC_SAMPLES	7 23 39 46	95	45	45	100

Figure 6.10: Performances of classification 1, discriminating PD vs PA for each microphone position and a number of feature equal to 4.

Considering classification 1 for four features, the best performances have been obtained in Position 2, 4 and Reference Microphone regarding the Fraction of Classified and the Pessimistic Accuracy.

Considering the Standard Accuracy, the best performances have been obtained in Position 2, with the iPhone 8 and MC Samples, while the optimistic Accuracy has reached values up to 100% in every position.

The features selected by the classifier are more dishomogeneous than those in the previous classification and furthermore there is a reduction of the percentage of the Fraction of Classified and of the Pessimistic Accuracy.

6.3.2 Classification 2

The Table in Figure 6.11 shows the best performances obtained for the second classification method for a number of feature equal to 2.

CLASSIFICATION 2: PD VS PA - 2 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1:4	23 37	95	60	55	91.67

Figure 6.11: Performances of classification 2, discriminating PD vs PA for a number of feature equal to 2.

Considering classification 2 for two features, the features and the percentages obtained for the various parameters are the same with respect to those of classification 1 for two features.

Table in Figure 6.12 shows the best performances obtained for the second classification method for a number of feature equal to 3.

CLASSIFICATION 2: PD VS PA - 3 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1:4	23 39 40	95	75	75	100

Figure 6.12: Performances of classification 2, discriminating PD vs PA and a number of feature equal to 3.

Considering classification 2 for three features, there is an increasing of the percentage of the parameters with respect to classification 2 for two features and to classification 1 for three features. Only the percentage of Standard Accuracy hasn't changed.

The Table in Figure 6.13 shows the best performances obtained considering the second classification method considering a number of feature equal to 4.

CLASSIFICATION 2: PD VS PA - 4 Features					
ARTIFICIAL VOWEL	FEATURES	STANDARD ACCURACY %	FRACTION OF CLASSIFIED %	PESSIMISTIC ACCURACY %	OPTIMISTIC ACCURACY %
POSITION 1:4	2 6 41 42	75	45	45	100

Figure 6.13: Performances of classification 2, discriminating PD vs PA for each microphone position and a number of feature equal to 4.

Considering classification 2 for four features, there is a reduction of the percentages of the Standard Accuracy, Fraction of Classified and Pessimistic Accuracy with respect to classification 2 for two features and to classification 1 for three features, while the percentage of the Optimistic Accuracy has reached values up to 100% even in this case.

Chapter 7

Conclusions

In this thesis work a classification method has been proposed based on the estimation of the extraction error of the features to discriminate healthy subjects (HE) from patients with Parkinson's disease (PD) and pathological non-parkinsonian patients (PA) from Parkinsonian patients (PD).

In the first part of the work, from the chosen balanced dataset, consisting of 3 repetition of the sustained vowel '/a/' recorded with a microphone in air for the 10 subjects belonging to the HE, PA and PD classes, 30 new artificial vowels for each subject with the Monte Carlo sampling method have been generated.

The second part of the thesis work has been focused on the reproducibility and repeatability of the measurements.

The artificial vowels obtained previously have been reproduced through an "Head and torso simulator" in an anechoic chamber.

The vocal signals produced have been acquired using 3 different measurement chains:
- A microphone in air placed in 4 different positions;
- A reference microphone;
- A microphone embedded in an iPhone 8.

The purpose of such recordings has been to estimate the error of the feature extraction algorithm by comparing the sequences of the parameters generated by the Monte Carlo sampling method with those obtained from the recordings in the anechoic chamber.

The results highlight how the measurements showed no significant differences between estimated relative errors, suggesting that the measurements made can be considered reproducible and repeatable in different conditions.

The aim of the last part of the thesis work, has been focused on the training of a weighted classification algorithm to discriminate HE subjects from PD patients and PA patients from PD patients.

The combinations of features considered by the classifier have been selected in order to give more weight to subjects whose features have been extracted with less errors. The weights, in fact, have been defined as the reciprocal of the errors.

Cosidering the features extracted from the real signals, together with error estimates, the classification method provided the probability of belonging to HE class and to PA class.

The features extracted from the real signal, the artificial signal, and the signals obtained in the anechoic chamber, together with the measurement error, have been used to obtain two different subsets for the classification.

The real vowels and the artificial ones have been combined to perform two types of classification using a limited number of features (from 2 to 4).

A new metric for evaluating the accuracy of the classification model has been proposed.
Such metric is based on the classification rate of the subjects which excludes the non-classified subjects from the accuracy estimation.
Three parameters have been defined:

- Pessimistic accuracy;
- Optimistic accuracy;
- Fraction of classified;

Discriminating PD patients from HE patients, the best performances have been obtained in classification 1 (3 real vowels + 3 artificial vowels), for a number of features equal to 2, in Position 3, Position 4 and with the iPhone 8.
The Fraction of Classified and the Pessimistic Accuracy, have reached values up to 73.68%, while the Optimistic Accuracy has reached values up to 100%.

Discriminating PD patients from PA patients, the best performances have been obtained in classification 2 (3 real vowels + 12 artificial vowels), for a number of features equal to 3.

The Fraction of Classified has reached values up 75%, the Pessimistic Accuracy has reached values up 75% and the Optimistic Accuracy has reached values up to 100%

The classification showed some limitations due to the reduction of the dataset size. Such reduction highlighted a drop on the performances of the classifier regarding the fraction of classified and accuracy.

Considering the two classifications with 2, 3 and 4 features, the increasing number of features leads to a rise of the error on the classification.
This happens because the proposed uncertainty model is additive so the more feature are used, the more is the uncertainty.

This thesis work can be considered a first step of a study whose future objective is to improve the model under analysis, increasing the database of the monitored subjects, using devices such as mobile phones that allow the patient to record vocals following repeatable procedures, without the need to go to health facilities and without impairing their daily activities.

AppendixA

Avevo un bulldog che si chiamava Bulka. Era tutto nero salvo una macchia bianca all'estremità delle zampe anteriori.
Nei cani di questa razza, la mandibola è sempre prominente, così i denti superiori vengono a collocarsi dietro a quelli inferiori.
Ma quella di Bulka era tanto grossa che tra gli uni e gli altri denti rimaneva molto spazio.
Aveva il muso largo, grandi occhi neri e brillanti e i canini sempre scoperti, perfettamente bianchi. Somigliava a un grugno.
Bulka era assai forte. E se afferrava qualcosa tra i denti non c'era verso che mollasse la sua preda.
Stretti i canini nella carne dell'avversario, serrava la mascella e rimaneva sospeso come un cencio ad un chiodo: attaccato come una sanguisuga.
Un giorno che era stato lanciato contro un orso, gli afferrò tra i denti un orecchio.
L'orso cercava di colpirlo con una zampa, scuoteva la testa, ma non se ne poteva sbarazzare: finì per rovesciare il testone in terra per schiacciarvi il cane.
Su quest'ultimo, però, perché lasciasse la presa, dovemmo gettare una secchia di acqua gelata. Lo avevo avuto da ragazzo e gli davo da mangiare io stesso.
Quando dovetti partire a prestar servizio ne Caucaso, decisi di non prenderlo con me e cercai di andarmene senza che lo sapesse. Ordinai che lo tenessero rinchiuso.
Ero giunto alla prima tappa, stavo per ripartire con i cavalli freschi, quando ad un tratto notai una palla nera e brillante che avanzava velocissima sulla strada.
Era Bulka col suo collare di rame al collo. Correva a perdifiato; si gettò su di me, mi leccò la mano e poi, la lingua ciondoloni, si stese all'ombra sotto la vettura.
Seppi più tardi che aveva rotto un vetro per seguirmi; era saltato dalla finestra: aveva percorso venti chilometri d'estate, sotto un sole bruciante.

Bibliography

[1] G. E. MD, "Parkinson's, alzheimer's, and the new science of hope," *Psychology Today*, pp. 183–195, 2019.

[2] P. R. Tiwari PC, "The potential role of neuroinflammation and transcription factors in parkinson disease," pp. 71–80, 2017.

[3] e. a. Christopher G. Goetz, Barbara C. Tilley, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): Scale presentation and clinimetric testing results," pp. 2129–2170, 2008.

[4] B. . Kjæ, "Technical documentation : Head and torso simulator (hats) type 4128-c," 2010.

[5] C. Fan, Z. Xie, Y. Liu, C. Li, and H. Liu, "Adaptive controller based on spatial disturbance observer in a microgravity environment," *Sensors*, vol. 19, p. 4759, 11 2019.

[6] F. K. I. Cantürk, "A machine learning system for the diagnosis of parkinson's disease from speech signals and its application to multiple speech signal types," pp. 41:5049–5059, 2016.

[7] ""diagnosi e terapia della malattia di parkinson e parkinsonismi"." [Online].

[8] C. Klein and A. Westenberger, "Genetics of parkinson's disease," 2012.

[9] ""malattia di parkinson e parkinsonismi"." [Online].

[10] ""rating scales"," 2017. [Online].

[11] F. S. G. M. C. D. S. Fahn e R. Elton, "Recent developments in parkinson's disease," 1987.

[12] ""gondola parkinson"." [Online].

[13] W. D. Colle, "Voce computer,analisi acustica digitale del segnale verbale (il sistema csl-mdvp)," Omega Edizioni, 2001.

[14] P. V. C. N. R. e. M. D. B. Y. Maryn, P. Corthals, "«toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels,»," pp. 540–555, Journal of Voice, 2010.

[15] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound.," pp. 97–110, 1993.

[16] G. E. P. Antonella Castellana, Federico Casassa, "Nuovi parametri acustici utili nella diagnostica e nella prevenzione di patologie vocali," 2015.

[17] D. D. M. e. G. S. G. Y. D. Heman-Ackah, "«the relationship between cepstral peak prominence and selected parameters of dysphonia,»," pp. 20–27, 2002.
[18] e. a. Yolanda D. Heman-Ackah, Robert T. Sataloff, "Quantifying the cepstral peak prominence, a measure of dysphonia," 2014.
[19] J. S. H. R. J. Hillenbrand, A. Houde, "Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech," pp. 311–321, 1996.
[20] Y. A. Shreider, "The monte carlo method: The method of statistical trials," 1996.
[21] L. Martino, "A review of multiple try mcmc algorithms for signal processing," pp. 134–152, 2018.
[22] B. . Kjæ, "Head and torso simulator types 4128-c and 4128-d handset positioner for hats type 4606," 2010.
[23] ""alpine-mrp-f200"." [Online].
[24] ""motu audio express"." [Online].
[25] E. K. E. C. Blessie, "A feature selection algorithm using correlation based method, journal of algorithms computational technology," 2012.
[26] H. L. Lei Yu, "Feature selection for high-dimensional data: A fast correlation-based filter solution, conference: Machine learning, proceedings of the twentieth international conference (icml 2003)," 2003.
[27] E. W. A. Ly, M. Marsman, "Analytic posteriors for pearson's correlation coefficient, statistica neerlandica," 2015.
[28] A. C. Leon, "Comprehensive clinical psychology," pp. 243–285, 1998.