Politecnico Di Torino

Master Degree in Biomedical Engineering

Post-surgery monitoring of the phonatory system through the analysis of voice signals

Comparison between microphone in air and contact microphones

Supervisor

Prof. Alessio Carullo

Candidate

Viviana Cennamo

S232545

Co-Supervisors

Prof. Alberto Vallan

Eng. Antonella Castellana

Academic Year 2017 - 2018

Alla mia famiglia

Abstract

The diagnosis of voice pathologies currently requires endoscopy procedures, such as videostroboscopy, that allows the anatomy of the larynx and the vocal folds to be observed. Although this examination is the most important for instrumental voice assessment, it is at the same time intrusive, only provides instantaneous vocal-health status and it is performed only in clinics.

The investigation of the voice quality through the parametric analysis of vocal signals is an alternative technique that has been spreading during the last decades. It represents a non-invasive method that could help physicians in the diagnosis of pathologies related to the vocal apparatus.

The main goal of this thesis consists in providing an objective support in vocal folds surgery rehabilitation through the monitoring of the vocal signal and the estimation of voice-related parameters. The study has involved 8 patients at Molinette Hospital (Turin) affected by different vocal diseases. In particular, the patients were subjected to three acquisition sessions at three different times: the first just before the operation, the second and the third ones 1 month and 2 months after the operation, respectively. The data were simultaneously collected with a microphone in air and two contact microphones, a piezoelectric device and a Electret Condenser Microphone (ECM). The monitoring protocol was based on the acquisition of different speech materials: three repeated sustained vowel /a/, a reading task and a free speech task. The 8 participants also filled in a voice self-assessment questionnaire before the

surgical intervention and after two months.
The acquired signals have been processed in order to extract frequency perturbation parameters, amplitude perturbation parameters, and Harmonic-to-Noise Ratio only for the vowels /a/. In addition, the Cepstral Peak Prominence Smoothed (CPPS) descriptive statistics have been estimated both for vowels /a/ and reading and free-speech tasks. In particular, for vowel /a/, a procedure has been implemented with the aim to identify a minimum subset of parameters based on a correlation analysis. The identified subset includes two parameters related to frequency perturbation (local Jitter - Jitt - and Pitch Period Perturbation Quotient - PPQ), two parameters related to amplitude perturbation (local Shimmer - Shim - and Peak Amplitude Variation - vAm), the Harmonic-to-Noise Ratio (HNR), and the 5^{th} percentile and the standard deviation of the CPPS distributions. The same subset of parameters have been evaluated for vowels /a/ acquired with the three microphones. For continuous speech, the investigated parameters are the 95^{th} percentile (for microphone in air), standard deviation (for the three microphones) and mean (for the contact microphones) of CPPS distributions. This choice is based on results that exists in literature.
A preliminary validation step has been performed in order to compare the parameters estimated through specifically developed Matlab scripts to a commercial software environment, which is the Multi-Dimensional Voice Program (MDVP). This validation has involved both pathological and healthy subjects and has been performed for vowels /a/ acquired with the microphone in air. The obtained results have shown an high correlation for the parameters Shim and vAm and a lower correlation for the parameters Jitt and PPQ.
In order to evaluate the discrimination power between pre-surgery and post-surgery voice, the Mann-Whitney U-test was performed for the microphone in air and the piezoelectric contact microphone. The best parameter detected was

the PPQ both for microphone in air and contact microphone. For continuous speech in most cases the CPPS distributions obtained before operation have shown higher occurrences at lower values, since for pathological voice most speech becomes noisy sounds. The same subjects CPPS distribution, after the operation, showed a longer left tail and $CPPS_{\mathrm{mean}}$ values higher than those before operation.

All the parameters of the contact microphones were compared to the parameters of microphone in air. The Pearson's correlation coefficient has provided high values for the piezoelectric contact microphone especially for CPPS parameters. The positive results obtained with the contact microphone are forward-looking to extend this study to everyday life monitorings. Certainly the achieved results need to be supported by an increase of data-set, since during the research was difficult to find available and suitable patients.

In the last part of this thesis a statistical analysis for discrimination of healthy and pathological voices in sustained vowel /a/ was carried out. The study involved a group of 40 subjects (20 unhealthy and 20 healthy) which are part of already collected data-base. A single-variable regression model was firstly performed with one CPPS parameter at a time and in addition also the parameter Jitt, PPQ, Shim and HNR were considered. Then, a double-variable regression model was performed with the parameters that individually showed the best performance. The model which uses $CPPS_{\mathrm{5prc}}$ and PPQ as independent variables has been the most powerful in discrimination between the two categories. Eventually, a logistic regression with $CPPS_{\mathrm{5prc}}$,PPQ and Shim was performed but the results have not improved, so it is not suggested to complicate the model.

Contents

Chapter 1

Introduction

This chapter deals with a description of the physiology of vocal apparatus and phonation production. Then the common vocal pathologies and their effect on vocal apparatus are reported. In the end the attention has been fixed on the state of the art choice of vocal parameters.

1.1 Vocal apparatus and phonation

The vocal apparatus includes the lungs in order to supply air pressure and airflow, the vocal folds whose vibration modulates the airflow and produces voice source, and the vocal tract that modifies the voice source. The airspace between the two opposing vocal folds is the glottis. the cartilages of the larynx form a structure that supports and houses the vocal folds. When differentially contracted, the laryngeal musculature can move the cartilages relative to one another in order to open and close the glottis causing abduction and adduction [1]. The phonation is the process by which the vocal folds produce certain sounds through quasi-periodic vibration, it occurs when air is expelled from the lungs through the glottis, creating a pressure drop across the larynx. When this drop becomes sufficiently large, the vocal folds start to oscillate creating a vocal signal.

1.2 Vocal fold pathologies and their effect on phonation

Age, smoke, type of job associated with vocal overuse (teacher, singer, etc.), environmental conditions are some of the many causes of vocal fold lesions. These can impact at not only physical but also social, emotional and economica level, restricting the participation to normal activities of life. Below are reported some of the most common vocal diseases.

1.2.1 Polyps

Vocal cords polyp is a benign formations that develop on the vocal cords, that hinder the correct vibration and the complete adduction during the emission of the sounds. They can be caused by repeated stress of the vocal cords which in the long run lead to the formation of chronic inflammatory states.
The vocal cords are anatomical structures poor in pain receptors and therefore, contrary to other organs of the body, even if they are inflamed, the pain signal is missing. It is therefore possible that unknowingly vocal cords are subjected to considerable efforts while they are suffering and inflamed. This compromises the ability to keep voice output constant and leads to changes that can affect the amplitude of vibration without affecting the frequency. The typical symptom associated with the formation of a polyp of the vocal cords is that of hoarseness. In this case it is necessary to undergo a laryngoscopy, so as to verify the state of the laryngeal structures and in particular that of the cordial mucosa. If the treatment based on drugs and voice rest should be unsuccessful, since it is still a persistent neoformation, is necessary to remove

the lesion through laryngeal microsurgery to restore a good quality of the voice.

1.2.2 Nodules

They represent one of the most common benign lesions of the vocal cords and are a typical consequence of the misuse of the voice. From the histological point of view they are located at the level of the cordial mucosa and appear as whitish thickenings sometimes affecting only one vocal cord but more often than both (assuming in this case the typical definition of "kissing nodules"). Normally they are observed at the level of the middle-anterior third of the free cordal edge, since this area is the one most exposed to the stresses of phonation. Very often the nodules are manifested in subjects who for professional reasons make extensive use of the voice, such as teachers and singers. From the mechanical point of view, due to the reduction of plasticity of the cordial epithelium, an incomplete closure of the glottis occurs. From the symptomatic point of view this translates into a hoarse, blown and easily tiring voice.

1.2.3 Cysts

Cysts are enclosed, sac-like structures that are typically of a yellow or white colour. The symptoms of vocal fold cysts commonly include a hoarse voice and problems with the pitch of the voice. Initial treatment of the cysts involves voice therapy to reduce harmful vocal behaviours. If symptoms remain, patients may require surgery to remove the cyst.

1.2.4 Laryngeal nerve paralysis

The motor innervation of the larynx is provided by the vagus nerve through the external branch of the superior laryngeal nerve and the inferior, or recurrent, laryngeal nerve (which innervates all other laryngeal intrinsic muscles).

Between the two laryngeal nerves (upper and lower), the paralysis of the inferior is by far the most frequent and most relevant from the clinical point of view. With the exception of rare forms related to neurological or tumor-related pathologies, the main causes of recurrent nerve paralysis are represented by surgical traumas resulting from various types of operations performed at the neck level. The deficit of addictive motility of the vocal cord and the consequent glottic insufficiency are aggravated over time by the progressive atrophy of the paralyzed cord. The functional consequence of the vocal fold paralysis is represented by a clinically very evident impoverishment of the phonatory function, with a raucous voice, sometimes with raising of the tonality, frequent appearance of falsetto and reduction of phonation duration.

1.2.5 Sulcus vocalis

It is an area of whitish cut at the level of the free cord. This groove causes vocal symptoms such as: phonatory fatigue, veiled vocal timbre and poor of harmonics. The voice of patients suffering from sulcus has a vailed tone and is poor in harmonics.

1.3 The acoustic characteristics of the vocal signal

The closing and opening frequency of the glottis determines the fundamental frequency (f0), or first harmonic, of the generated signal. In normal conditions the glottis signal is defined as complex and quasi-periodic, because the frequency and amplitude characteristics can change over time [2]. These modifications can occur in a short period of time and are indispensable to give naturalness to speech. A necessary condition for the glottis source to produce a periodic signal is the complete contact of the vocal cords. If this does not

happen due to pathological or voluntary causes, the generated signal will not be periodic but aperiodic, for instance this cause brethiness of voice.
The f0 average it is peculiar to every individual and varies according to age, sex, type of vocal activity. we can classify the fundamental frequency average in different ranges:

- 255-440 Hz for kids;
- 175-245 Hz for women;
- 105-160 Hz for men;

The fundamental frequency of the glottal signal remains the same also after passing through the vocal tract, where it became the vocal signal [2]. The glottal signal has a relatively simple waveform compared to the vocal signal as we can see from figure 1.1. The first is recorded with a contact microphone and the second with a microphone in air that considers also the vocal tract.

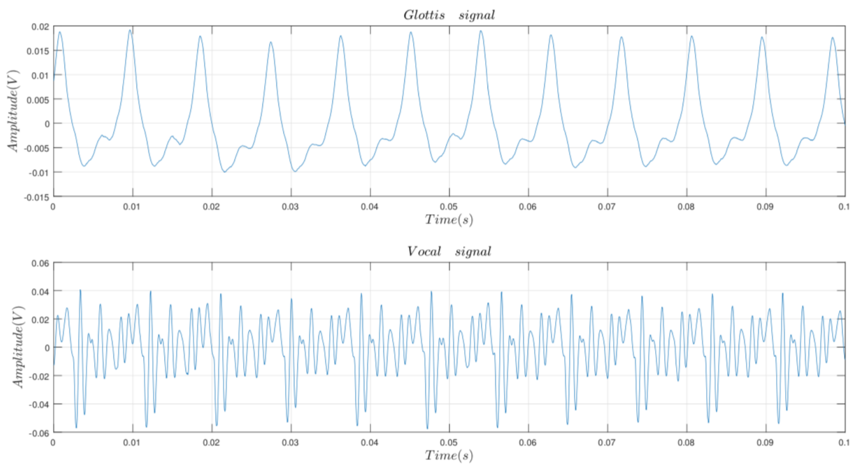

Figure 1.1: The glottis signal is shown in the upper part of the figure and the vocal signal is shown at the bottom down, for sustained vowel /a/.

Chapter 2

State of the art in vocal analysis

The physicians use different instruments to recognize and to diagnose vocal disorders. The most used is the videostroboscopy, a video recording that uses a flexible fiberscope. This technique allows the observation of the anatomy of the larynx and the vocal folds, to give an evaluation of the vocal disorder from an anatomic and organic point of view.

2.1 Qualitative analysis of voice

Perceptual evaluation of the voice is still considered the fundamental investigation in the clinical and instrumental evaluation of the voice and can not be completely replaced by instrumental measures. However, to be valid it must be done with standard procedures directly listening to the patient's voice by speech therapists. One of these approaches is the GRBAS scale, or GIRBAS scale, proposed by Hirano (1981), which became the standard scale for speech therapists. Every letter of the acronym GIRBAS refers to a qualitative characteristic of voice (tab. 2.1) and for each of them a number from 0 to 3 is associated. The number 0 is associated to an healthy voice, instead the number 3 refers to a seriously unhealthy voice.

GIRBAS scale	
Component	**Description**
G - Grade	General grade of dysphonia
I - Instability	Changes in voice quality over time
R- Roughness	Impression of irregularity of the vibration of the vocal folds
B - Breathiness	Degree of breath voice
A - Asthenia	Degree of weak voice
S - Strain	Degree of strain and hyperfunctional use of phonation

Table 2.1: Girbas scale description. Rating scale:0, normal; 1, slight; 2, moderate; 3, severe

2.2 Multiparametric analysis of the vocal signal

The only audition judgment is incomplete and less reliable, that's why the perceptual approach is commonly supported by an objective approach that uses signal processing techniques for measuring acoustic features in the time and frequency domain.

In the last years, acoustic measurement have taken on a fundamental importance, because they are non-invasive, relatively low cost, real-time and easy of application [3].

Typical features are *fundamental frequency* (f_0), perturbation measures such as *jitter* (changes in pitch with time) and *shimmer* (changes in amplitude with time), and *harmonics-to-noise ratio* (HNR). These parameters can be obtained only in continuous vowels [4],[5].

The fundamental frequency f_0 is not used in disorders estimation because it's affected by several factors like age, gender, lifestyle and job. The jitter rappresents the perturbation of frequency from cycle to cycle linked to a lack of control of the vocal folds vibration. Insted, the shimmer is the corresponding in amplitude of the jitter and it's correlated to noise and breathiness [4]. The HNR represents the ratio between the energy of harmonic and the noise energy.

In the last few years, these parameters were used also in voice pathology detection. In 2006 P. Gomez-Vilda et al. showed that perturbation mesaures next to bimechanical parameters increases the accuracy of acoustic analysis, improving the detection of voice pathology, like polyps, nodules and Reinke's edema [6]. According to Nicastri et al. [7] The amplitude parameters are the best in discrimination of vocal disorders like polyps and cysts, because the lack of complete closure of the glottis creates a breath which compromises the ability to produce a constant sound emission and leads to changes which can affect the vibratory amplitude without affecting the frequency.

The mentioned parameters has the limitation that are defined only in sustained vowels and are very linked to f_0, so small errors can change a lot the results [8].

2.3 Spectral and Cepstral-based measures

In order to bypass the already mentioned problems and to give a much complete parametric analysis also in continuos speech, several studies have turned towards approaches in frequency domain. In particular, spectral and cepstral measures do not require cycle boundary detection, therefore they are more accurate for dysphonic voices. The power spectral estimation method allows to obtain important information about the voice but the research is moving toward a new important approach based on cepstral mesaurement.

In 1963 Borget et al. [8] defined the concept of cepstrum as the spectrum of a spectrum, specifically "the log-power spectrum of the log-power spectrum of a

signal". So the cepstrum is defined in the time domain so the terms *cepstrum* and *quefrency* are the corresponding of the words spectrum and frequency. In 1994 Hillenbrand et al. were the first who considered this representation useful in predicting breathiness [9].
In particular, the cepstrum shows the peak of energy of every frequency component. The periodic signals show a defined harmonic structure and the cepstrum contains a visible peak corresponding to fundamental period (fig.2.1). The amplitude of the peak defines the parameter CP (Cepstral Peak) and it's

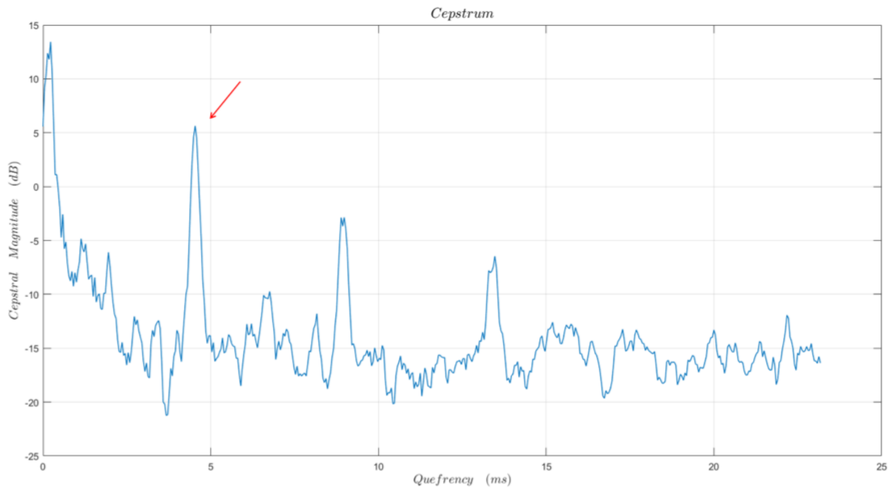

Figure 2.1: Example of Cepstrum.

located approximately between 3 ms and 16 ms (60 Hz and 300 Hz). This parameter is not affected only by the harmonic organization but also by the overall energy of the signal and consequently influenced by noise. So in the same work Hillenbrand defined the parameter CPP (Cepstral Peak Prominence), which is the difference between the pick and the corresponding value at the same quefrency of the linear regression. After two years he optimized his work adding a smoothing in order to have a more considerable peak, so the parameter CPPS is defined (Cepstral Peak Prominence Smoothed) [10]. A description is offered in fig. 2.2. The CPPS distribution is built considering

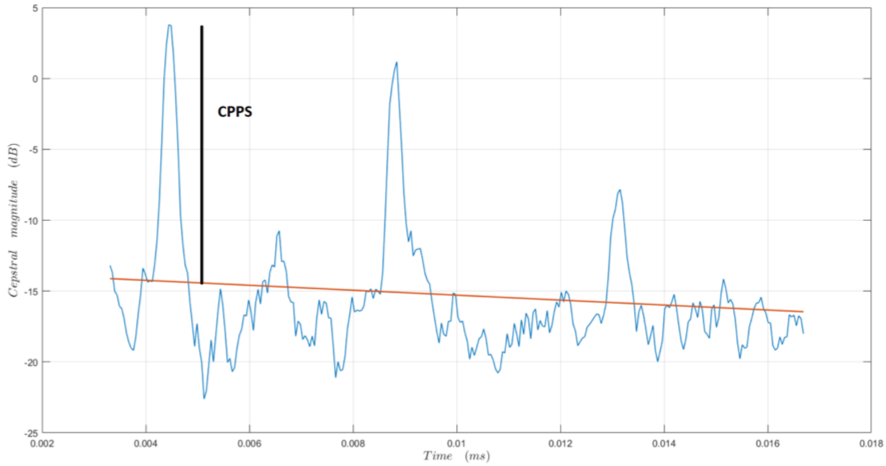

Figure 2.2: Example of Cepstral Peak Prominence smoothed with regression line.

different frames shifting a window along the signal with overlap. So the result is composed by different CPPS value and it's possible to extract different statistics, the first was the mean. Few years later in 2002, other researches have continued to study breathiness and roughess and other statistics were considered, like median, mode, range and standard deviation of CPPS [11]. The wide presence of CPPS studies allowed the born of different software, like praat, speech tool and ADSV, to calculate the statistic parameters. As a consequence in 2014 Maryn et al. in 2014 compared the different results of praat and speech tool [12] and in 2016 Watts et al. compared Praat and ADSV. In 2017 Castellana et al. have studied different statistics for CPPS distribution of sustained vowel /a/ in order to discriminate healthy and unhealthy subjects. The acquisition were made not only with microphone in air but also with a contact microphone [13]. In 2018, in another study conducted by Castellana et al. the a specific developed Matlab script was used to compare the results of CPPS mean with the ones gave by Hillenbrand algorithm. The comparison has indicate a robust linear correlation between the two algorithms. Different descriptive statistics of CPPS distribution have been extracted with the aim to find the best discriminant between healthy and unhealthy voices. The results have shown the fifth percentile as the best discriminator for microphone in air and the standard deviation for contact microphone in sustained vowel /a/ [14], instead in another study the ninety-fifth percentile has shown the best results in continuous speech [15].

Chapter 3

Post-surgery monitoring

In this chapter is described how the post-surgery monitoring has been carried out. Firstly, the group of subjects involved in the study and how the data were collected are described. Than the procedures of pre-processing, processing methods and instruments to achieve the results were illustrated. Finally, the results relative to vowel /a/ and continuous speech for each microphone are discussed.

3.1 Choice of subjects

The subjects involved in this study are 8 Italian patients at Molinette Hospital that in time between March and July of 2018 underwent surgical treatment of the vocal folds. This statistic sample consists, particularly, of 7 females and 1 male between 29 and and 82 years old affected by different vocal fold lesions. In table 3.1 there is a summary of the subjects involved in the study.

Subject	*Age*	*Gender*	*Pathology*
1	74	F	**Sulcus vocalis** of the left vocal cord
2	71	F	**Reinke's edema** of the left vocal cord
3	29	F	**Polyp** of the right vocal cord
4	56	F	**Polyp** of the right vocal cord
5	82	M	**Paralysis** of the left vocal cord
6	59	F	**Cyst** of the left vocal cord
7	43	F	**Cyst** and **Edema** of the left vocal cord
8	39	F	**Cyst** of left vocal cord and **Edema** of the right one

Table 3.1: Summary table of subjects involved into the study.

3.2 Data collection and protocol

The procedure started with the recording of the voice of the patients three times: the fist just before the operation, the second and the third 1 month and 2 months after the operation respectively. In all the tests the patients concurrently wore three microphones (fig. 3.2), two contact and one in air, in order to have three different recordings for each patient.

After activating the microphones each patient went into a quiet chamber and was asked to follow an established protocol:

1. Vocalize the vowel /a/ three times, on a comfortable pith and loudness. The vowel is maintained for as long as possible from 3 s to 10 s. Between a vowel and the next the subject can wait for the necessary time to catch his breath;

2. Read a phonetically balanced Italian passage (Appendix A), without interruption from the beginning to the end;

3. Speech for 1 minute, without stop, about a free topic;

NOME: COGNOME:

ETA’: SESSO: M F

FUMATORE: SI NO NOTE:

PROFILO DI ATTIVITÀ E PARTECIPAZIONE VOCALE - PAPV
Fava, Paolillo, Oliveira, Behlau

Ti invitiamo a rispondere ad ogni domanda mettendo una croce (X) su qualsiasi punto della linea che rappresenta al meglio il grado della tua risposta. Una croce sull’estremo lato sinistro della linea indica che il problema non è MAI presente; una croce sull’estremo lato destro della linea indica che il problema è GRAVE o è SEMPRE presente; una croce su qualsiasi punto della linea tra i due estremi, andando da sinistra a destra, indica che il problema è gradualmente più grave o più frequente

Autopercezione dell’*entità* del problema vocale

1 Attualmente qual è l’entità del tuo problema vocale?
Lieve ________ Grave

Effetti sul lavoro

2 **Il tuo lavoro risente del tuo problema vocale?**
Mai ________ Sempre

3 Negli ultimi 6 mesi hai pensato di cambiare lavoro a causa dei tuoi problemi vocali?
Mai ________ Sempre

4 Il tuo problema vocale ha creato condizioni di stress sul tuo lavoro?
Mai ________ Sempre

5 Negli ultimi 6 mesi il tuo problema vocale ha influito sulle decisioni legate al futuro della tua carriera?
Mai ________ Sempre

Effetti sulla comunicazione quotidiana

6 A causa del tuo problema vocale la gente ti chiede di ripetere ciò che hai appena detto?
Mai ________ Sempre

7 Negli ultimi 6 mesi hai mai evitato di parlare con gli altri a causa del tuo problema vocale?
Mai ________ Sempre

8 La gente ha difficoltà a capirti al telefono a causa del tuo problema vocale?
Mai ________ Sempre

9 Negli ultimi 6 mesi hai ridotto l’uso del telefono a causa del tuo problema vocale?
Mai ________ Sempre

10 Il tuo problema vocale influenza il tuo modo di comunicare in ambienti silenziosi?
Mai ________ Sempre

11 Negli ultimi 6 mesi hai mai evitato conversazioni in ambienti silenziosi a causa del tuo problema vocale?
Mai ________ Sempre

12 Il tuo problema vocale influenza il tuo modo di comunicare in ambienti rumorosi?
Mai ________ Sempre

13 Negli ultimi 6 mesi hai mai evitato conversazioni in ambienti rumorosi a causa del tuo problema vocale?
Mai ________ Sempre

14 Il tuo problema vocale influisce su ciò che vuoi comunicare quando parli a un gruppo di persone?
Mai ________ Sempre

15 Negli ultimi 6 mesi hai mai evitato conversazioni di gruppo a causa del tuo problema vocale?
Mai ________ Sempre

16 Il tuo problema vocale ti impedisce di far capire quello che vuoi comunicare?
Mai ________ Sempre

17 Negli ultimi 6 mesi hai mai evitato di parlare a causa del tuo problema vocale?
Mai ________ Sempre

Effetti sulla comunicazione sociale

18 Il tuo problema vocale influisce sulle tue attività sociali?
Mai ________ Sempre

19 Negli ultimi 6 mesi hai mai evitato attività sociali a causa del tuo problema vocale?
Mai ________ Sempre

20 I tuoi familiari, amici o colleghi di lavoro sono infastiditi dal tuo problema vocale?
Mai ________ Sempre

21 Negli ultimi 6 mesi hai mai evitato di comunicare con i tuoi familiari, amici o colleghi di lavoro a causa del tuo problema vocale?
Mai ________ Sempre

Effetti sulle tue emozioni

22 Sei infastidito dal tuo problema vocale?
Mai ________ Sempre

23 Ti vergogni del tuo problema vocale?
Mai ________ Sempre

24 Hai poca stima di te stesso a causa del tuo problema vocale?
Mai ________ Sempre

25 Sei preoccupato per il tuo problema vocale?
Mai ________ Sempre

26 Ti senti insoddisfatto a causa del tuo problema vocale?
Mai ________ Sempre

27 Il tuo problema vocale influisce sulla tua personalità?
Mai ________ Sempre

28 Il tuo problema vocale incide sulla tua immagine?
Mai ________ Sempre

DATA:

Figure 3.1: Evaluation questionnaire PAPV.

Figure 3.2: Patient wearing the three microphones. In the background there is the recording chamber.

After the recordings, only before and 2 months after the operation, the patients have filled in a questionnaire, called “ Profilo di Attivitá e Partecipazione Vocale” (fig .3.1), that is composed by 28 questions about the perceptual effect of the vocal disorder in patient’s daily life like job and social activity [16].

3.3 Recording equipment

For the data collection, three different microphone were used: a microphone in air and two contact microphones. The first records the vocal signal after the vocal tract, consequently the obtained signal is more complex and affected by noise, but the recording is clearer to the ear than the other two contact microphones that capture the vibration of the vocal folds.
The sensors are:

- An omni-directional ***headworn microphone*** MIPRO MU-55HN (fig.

3.3). The microphone presents a flatness of $\pm 3dB$ in the range from 40 Hz to 20 kHz. It's connected to a bodypack trasmitter ACT-30T, which trasmits to a wireless system Mipro ACT 311. A recorder ZOOM H1 (Zoom Corp., Tokyo, Japan) records the output signal of system in a SD card with a sample rate of 44100 Hz and 16 bit of resolution. The microphone is placed at a distance of about 2,5 cm from the lips of the talker.

Figure 3.3: Headworn microphone (MIPRO) with the trasmitter.

- A contact ***Electret Condenser Microphone*** (ECM AE38, Alan Electronics GMbH (Dreieich, Germany)). It is positioned on the jugular notch and it is fixed by means of a surgical band in order to catch the vocal folds vibration through the skin. A recorder ROLAND R05 (Roland Corp., Milano, Italy) (fig. 3.4) records the signal sampled at 44100 Hz with 16 bit of resolution and memorizes it in a SD Card.

- A contact ***Piezoelectric Contact Microphone***(HX-505-1-1, HKKK, 406, PLant 1, Jiadind Science Park, Dalang, Longhua New Dist., Shen-

Figure 3.4: Electret Condenser Microphone (ECM) connected to the recorder ROLAND R05.

zhen, Guangdong, China). It is a neck-ring positioned near the jugular notch that catches the vibration of vocal folds (fig.3.5). Through an AUX cable it is connected to a smartphone (Samsung SM-G310Hn). The "Vocal Holter App" downloaded to the smartphone records the signals with a sample rate of 22050 Hz and resolution of 16 bit.

3.4 Vocal Parameters

In this study, starting from the parameters definition provided by the Multi-Dimensional Voice Program [17], 5 vocal signal parameters related to frequency perturbation (*Jita, Jitt, RAP, PPQ,* vF_0) and 4 related to amplitude perturbation (*Shim, ShdB, APQ, vAm*) were evaluated. In figure 3.6 is possible to observe a schematic rapresentation of Jitter and Shimmer. Furthermore, 1 parameter, related to spectral energy balance between harmonic components and disharmonic components (*HNR*), was added into the study. In particular these were evaluted only for vowel /a/. For this reason, to make even the

Figure 3.5: Piezoelecrtic Contact Microphone (PIEZO) connected to the Samsung smartphone.

study of continuous speech possible, 9 parameters related to CPPS distribution were evaluated(*CPPSmean,CPPSmedian, CPPSmode, CPPSstd, CPPSrange, CPPS*5^{th}*percentile, CPPS* 95^{th} *percentile,CPPSskewness and CPPSkurtosis*).

3.4.1 Parameters related to period perturbations

After the extraction of fundamental Period (T_0) it's possible to evaluate the modification in time of this value. The parameters are:

- ***JITA***(μs) is the **absolute jitter** and describes the absolute mean variation period to period of the fundamental period T_0 (Ferrero et al.,1995 [20]):

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

  where $T_0^{(i)}$, with $i = 1, 2, ..., N$, are the periods extracted by the vocal signal and N is the number of periods. This parameter describes the mean of the difference between one period and the next one.

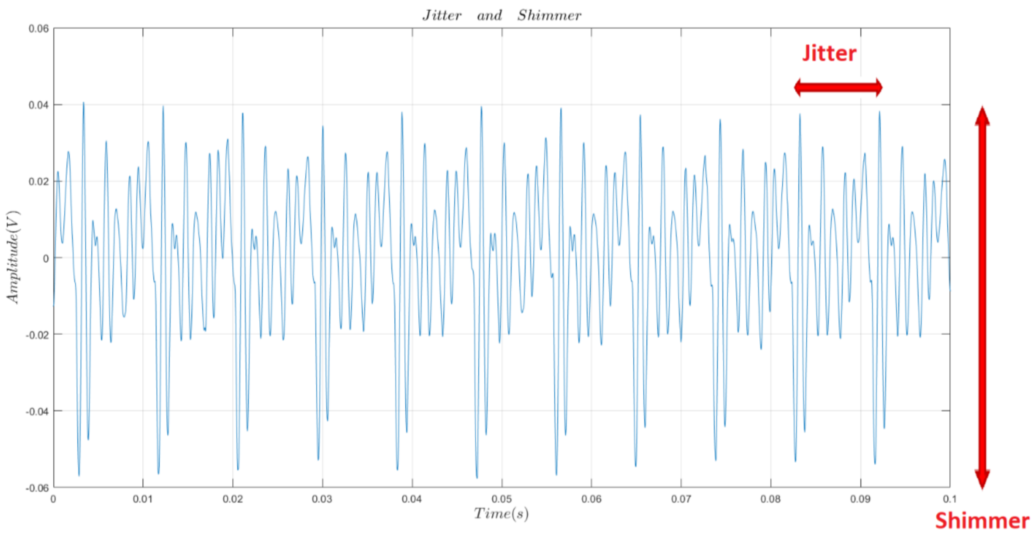

Figure 3.6: Representation of Jitter and Shimmer in vocal signal .

- ***JITT***(%) is the **local jitter** and describes the relative mean variation period to period of the fundamental period:

$$Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_0^{(i)} - T_0^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N}T_0^{(i)}}$$

where $T_0^{(i)}$, with $i = 1, 2, ..., N$, are the periods extracted by the vocal signal and N is the number of periods. The formula is similar to the Jita one, and it differs for the division of the Jita for the average fundamental period.

- ***RAP***(%) is the **Relative Average Perturbation** of 3 in 3 periods with the step of one of the fundamental period. The formula is:

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

where $T_0^{(i)}$, with $i = 1, 2, ..., N$, are the periods extracted by the vocal signal and N is the number of periods. The RAP is similar to Jitt but in this case, instead of calculating the difference between one period and the next, the average of three periods is calculated (3 smoothing factor) than is subtracted the value of the central period.

- ***PPQ***(%) is the **Pitch Period Perturbation Quotient** and gives the relative average perturbation of 5 in 5 periods (5 smoothing factor):

$$PPQ = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\left|\frac{T_0^{(i-1)}+T_0^{(i)}+T_0^{(i+1)}}{3} - T_0^{(i)}\right|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}}$$

where $T_0^{(i)}$, with $i = 1, 2, ..., N$, are the periods extracted by the vocal signal and N is the number of periods.

- ***vF*₀**(%) is the **Fundamental Frequency Variation**. It's the relative variability of standard deviation of F_0 with respect to the calculated mean fundamental frequency:

$$vF_0 = \frac{\sigma}{F_0} \times 100 = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_0 - F_0^{(i)})^2}}{\frac{1}{N}\sum_{i=1}^{N} f_0^{(i)}} \times 100$$

where F_0 is the average fundamental frequency, σ is the standard deviation of f_0, and $f_0^{(i)}$ are the individual frequency values extracted.

3.4.2 Parameters related to amplitude perturbations

The irregularity cycle after cycle of the amplitude may be due to the inability of the vocal cords to sustain a periodic vibration with a defined period. The parameters investigated are:

- ***ShdB***(dB) is the **absolute shimmer** that describes the absolute average variability period by period of the pick to pick amplitude:

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20log\frac{A^{(i+1)}}{A^{(i)}}\right|$$

where $A^{(i)}$, with i = 1,2,..,N, are the amplitude peak to peak and N is the number of impulses extracted.
The absolute shimmer is very sensitive to the amplitude variations occurring between consecutive pitch periods, so it gives a measure of the *short-term* amplitude perturbation.

- ***Shim***(%) is the **local Shimmer** and describes the relative evaluation of the period-to-period (very short term) variability of the peak-to-peak amplitude:
$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A^{(i)} - A^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}}$$
where $A^{(i)}$, with i = 1,2,..,N, are the amplitude peak to peak and N is the number of impulses extracted. Both Shim and ShbB are relative evaluations of the same kind of amplitude perturbation but they use different measures for the result, percent and dB.

- ***APQ***(%) is the **Amplitude Perturbation Quotient** and describes the relative variability of 11 to 11 periods (smoothing factor 11) with step of 1:
$$APQ = \frac{\frac{1}{N-10}\sum_{i=1}^{N-10}\left|\frac{1}{11}\sum_{r=0}^{10} A^{(i+r)} - A^{(i+5)}\right|}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}}$$
where $A^{(i)}$, with i = 1,2,..,N, are the amplitude peak to peak and N is the number of impulses extracted. APQ it is less sensitive to pitch extraction errors, than shim, bu it's still provides a reliable indication of short-term amplitude variability in the voice.

- ***vAm***(%) is the **Peak Amplitude Variation**. It gives relative variability of the peak-to-peak amplitude variations (short to long-term) within the analyzed voice sample:

$$vAm = \frac{\sigma}{A_0} \times 100 = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_0 - A^{(i)})^2}}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \times 100$$

where $A^{(i)}$, with i = 1,2,..,N, are the amplitude peak to peak, A_0 is the average value of the extracted peak-to-peak amplitude and N is the number of impulses extracted.

3.4.3 Harmonic to Noise Ratio

HNR(dB) is the Harmonic to Noise Ratio. It is a measure that quantifies the amount of additive noise in the voice signal. It is the ratio between the components of harmonic spectral energy and the components of disharmonic spectral energy:

$$HNR = 10log_{10}\frac{r(\tau_{max})}{1 - r(\tau_{max})}$$

where $r(\tau_{max})$ is the local maximum of the normalized autocorrelation function and according to Boersma [18] it represents the relative power of the periodic (or harmonic) component of the signal and and its complement represents the relative power of the noise component.

3.4.4 CPPS distribution

The Cepstral Peak Prominence Smoothed is a measure of cepstral peak amplitude normalized for overall amplitude, smoothed across time and across quefrency (Hillenbrand et al., 1996). Given the vocal signal y, it is possible to define:

$$yFFT = 20log\,|FFT(y)|$$

$$yFFT2 = 20log\,|FFT(yFFT)|$$

where yFFT2 is the cepstrum of signal y. Generally the cepstrum is calculated for different windows in which the signal y is divided, then a smoothing in time domain and a smoothing in frequency domain are performed in order to have a good prominence of peak. The CPPS value is the difference between the amplitude of peak and the corresponding value in the regression line calculated on cepstrum. In the and we have many CPPS value for each window and it is possible to build a distribution where 9 descriptive statistics can be calculated: *mean, median, mode, standard deviation, range,* 5^{th} *percentile,* 95^{th} *percentile, skewness and kurtosis.*

3.5 Data processing

3.5.1 Pre-processing

The recordings obtained in each monitoring from the three microphones were stored on the personal computer in .wav format via SD cards. Each recording was renamed and cut into five different audio files thanks to Audacity 2.2.2 software:

- *Vowel /a/ files*: are three for each recording (A1, A2 and A3) and around 10 seconds long, except for some subjects that were not able to maintain the vowel for long time. They were obtained cutting only the 4 central seconds of the vowels where it is more stable excluding the initial part and the final part;

- *Reading file*: was obtained cutting the reading part of the original audio file from the starting point to the word "sanguisuga" of the "Bulka" passage. In this way only the first nine sentences was analyzed reducing the computational time;

- *Free speech file*: was obtained cutting only 30 s of the 1 min recorded, for the same reasons mentioned above;

The signals of ECM and MIPRO were resampled at 22050 Hz while the PIEZO signals have been maintained at the original sampling rate (22050 Hz), in order to make possible the CPPS processing.

3.5.2 Vowel /a/ processing

After the pre-processing phase the vowel /a/ files were processed using two different Matlab ® R2017b algorithms:

1. The first is used in the analysis of amplitude and frequency perturbation parameters, and in harmonic to noise ratio parameter. The algorithm operates on one signal at time and compute an operation of autocorrelation, in which the maximum index is extracted in order to find the fundamental frequency expected of the vocal signal and the corresponding pitch period. Starting from this fundamental period (T_0) in samples, the signal position is moved ahead by the current T_0 in order to compute all the T_0 for jitter related parameters and all peak-to-peak amplitude for shimmer related parameters. The HNR parameter is obtained from the already mentioned formula (see paragraph 3.4.3) calculating the fundamental frequency for signal windows of 1024. The HNR values obtained in each window are mediated together to have a single value of HNR for each signal.

2. The second algorithm is used in CPPS distribution evaluation. The analysis is performed dividing the vowel signal into windows of 1024 samples (about 46 ms) with a 2 ms overlap. On each window the spectrum is calculated with the Matlab funcion FFT and is multiplied by the Hamming window, and the cepstrum is calculated operating a second FFT according to the formula. The cepstra obteined are time averaged with

a window of 14 ms and then they are quefrency averaged with 7 bin windows. According to the definition, the CPPS is calculated as the distance between the cepstrum maximum peak and the cepstrum regression line. The maximum search occurs between 3.3 ms and 16.7 ms, because the range of fundamental frequency in human voice is between 60 Hz and 300 Hz. The CPPS obtained for each frame are plotted through an Histogram and the 9 descriptive statistic parameters are calculated.

3.5.3 Reading and free speech processing

For reading and free speech files only CPPS parameters have been estimated. Before submit the signals to CPPS algorithm is necessary to remove the unvoiced segments, infact they are not explicative of the subject condition and lead to an alteration of CPPS distribution. For this purpose a silence removing algorithm have been implemented on software Matlab® R2017b. The aim of the algorithm is to find a threshold, adaptable to the signal, between voiced and unvoiced segments. The signal is divided in frames of 1024 samples and for each of them the value of Root-Mean-Square (RMS) is computed. The threshold is chosen as the 60% of the RMS of the whole signal on 1024-point windows. All the frames which RMS is higher than the threshold is considered voiced, those that do not meet the condition are discarded and considered unvoiced. This algorithm allows to adapt the threshold in relation with the amount of noise in a singular signal but surely it can be optimized. The files, saved again without the unvoiced parts, were submitted to the already mentioned CPPS algorithm (paragraph 3.5.2).

3.6 Results and Discussion: Sustained vowel /a/

3.6.1 Parameters correlation

After the data processing, all parameter of all vowel /a/ recordings were stored on a table, through Microsoft Excel. For each microphone a different table was compiled. As a preliminary phase, the parameters obtained were compared each other to find a minimum subset of uncorrelated parameters through the *Pearson's correlation coefficient* (ρ). This coefficient is a measure of the linear correlation between two variables X and Y:

$$\rho(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y}$$

where $cov(X, Y)$ is the covariance between X and Y, σ_X and σ_Y are the standard deviations of X and Y. The result is a value between $+1$ and -1, where 1 is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation.
In the table (fig.3.7) are reported the Person's correlation coefficient obtained by comparing each parameter of vowel /a/ with all the others one by one. The work was carried on for the vowels recorded with the microphones MIPRO and PIEZO obtaining two ρ values for each couple of parameters compared. The rad ones belongs to PIEZO vowel /a/ and the black ones to MIPRO vowel /a/. The boxes on the diagonal have been filled with black because they represent the comparison of each parameter with itself and this do not give information, furthermore only the part above the diagonal has been filled because it is mirrored to the one below. The green boxes identify the most correlated couples of parameters, with the highest ρ for each microphone. In particular the jita, and RAP parameters are very highly correlated to Jitt one with a value of $\rho = 0,99$ for both MIPRO and PIEZO microphones, also the vF_0 is

very correlated to Jitt with $\rho = 0,89$ for MIPRO and $\rho = 0,92$ for PIEZO. Furthermore, Shdb and APQ are highly correlated to Shim with $\rho = 0,99$ for both the microphones and Cpps95prc is correlated to Cpps5prc with $\rho = 0,81$ for MIPRO and $\rho = 0,85$ for PIEZO.

Accordind to this first result, the attention has been fixed on a minimum subset of parameters: *Jitt, PPQ, Shim, vAm, HNR, CPPS 95prc*(for continous speech), *CPPS 5prc* (for vowel /a/) and *CPPSstd*. After that, some vowel /a/ signals previously collected, belonging to both control and pathological subjects, were evalueted with Matlab algorithm and with MDVP software in order to compare the results.

***Pearson's correlation coefficient* (ρ) between vowel /a/ parameters**

vowMIPRO vowPIEZO	Jitt	Jita	RAP	PPQ	vF0	Shim	ShdB	APQ	vAm	HNR	CPPS 5prc	CPPS 95prc	CPPS Std
Jitt		0,99 0,99	0,99 0,99	0,69 0,52	0,89 0,92	0,89 0,74	0,88 0,72	0,88 0,73	0,42 0,25	-0,54 -0,64	-0,67 -0,52	-0,54 -0,57	0,67 0,46
Jita			0,99 0,99	0,71 0,52	0,89 0,91	0,89 0,78	0,89 0,76	0,9 0,77	0,42 0,25	-0,66 -0,63	-0,74 -0,54	-0,55 -0,43	0,66 0,48
RAP				0,69 0,52	0,89 0,92	0,89 0,73	0,88 0,71	0,89 0,72	0,42 0,24	-0,63 -0,59	-0,73 -0,5	-0,54 -0,4	0,66 0,45
PPQ					0,54 0,49	0,68 0,36	0,48 0,32	0,66 0,36	0,15 0,07	-0,47 -0,33	-0,53 -0,25	-0,56 -0,29	0,66 0,12
vF0						0,83 0,71	0,85 0,72	0,84 0,72	0,51 0,29	-0,69 -0,66	-0,76 -0,5	-0,49 -0,35	0,75 0,49
Shim							0,99 0,99	0,99 0,99	0,49 0,49	-0,81 -0,8	-0,76 -0,77	-0,63 -0,58	0,61 0,7
ShdB								0,98 0,99	0,49 0,55	-0,81 -0,81	-0,76 -0,78	-0,63 -0,58	0,67 0,72
APQ									0,49 0,5	-0,82 -0,82	-0,77 -0,79	-0,63 -0,72	0,63 0,72
vAm										-0,48 -0,32	-0,56 -0,55	-0,28 -0,55	0,64 0,35
HNR											0,74 0,72	0,55 0,45	-0,63 -0,77
CPPS 5prc												0,81 0,85	-0,8 -0,82
CPPS 95prc													-0,32 -0,4
CPPSstd													

Choice of a minimum subset of parameters:
Jitt, PPQ, Shim, vAm ,HNR, CPPS_5prc, CPPS_Std

Figure 3.7: Pearson's correlation coefficient between vowel /a/ parameters. For each couple of parameters compared there are two ρ values. The rad ones belongs to PIEZO vowel /a/ and the black ones to MIPRO vowel /a/.

3.6.2 Matlab Algorithm versus MDVP software

The MDVP system (Multi-Dimensional Voice Program) is a software which uses the hardware of the CSL-4300B (Computer Speech Laboratory). It provides a multidimensional analysis of voice with graphic and numerical presentation of analysis results. In particular, it returns 33 different parameters from a single voice segment [17] and for each of them a threshold for discrimination between healthy and unhealthy voices is provided. For our interest, the attention is focused on Jitt, PPQ, Shim and vAm. The MDVP requires a security key (which is available only to physicians) and accepts signals sampled at 44100 Hz and 16-bit quantization. In order to compare the results provided by the MDVP software and the matlab algorithm, 20 MIPRO vowels /a/ belonging to both controls and pathological were chosen. For this purpose, through Microsoft Excel, dispersion plots belonging to Jitt, PPQ, Shim and vAm are graphed (fig. 3.8, 3.9, 3.10, 3.11). Where, on x-axis there are the values returned by MDVP software and on y-axis the corresponding values provided by the Matlab script. A model of linear regression is executed and the R-square is extracted. The R^2 measures how far a linear model allows to approximate and predict the observed data. Values closed to 1 represent a good correlation between the two algorithms.

In particular the graphs show a very high correlation with Shim and vAm parametrs with $R^2 = 0,98$ and $R^2 = 0,99$ respectively, while they show a less linear correlation with jitt and PPQ with $R^2 = 0,52$ and $R^2 = 0,38$ respectively. This can be due to difference in the autocorrelation algorithm for the extraction of fundamental periods. Furthermore, some pathological signals are very irregular and it is difficult to extract the fundamental period correctly. This topic has already been treated in literature. In particular, Oguz et al. [19] have compared the acoustic parameters results obtained by MDVP and Praat. The latter is a free software used and supported by many clinicians all over the world, designed by Paul Boersma and David Weenink of the Phonetic Sciences Department of the University of Amsterdam. They showed that the parameters related to the fundamental period are not comparable between the two software, instead of the amplitude related parameters that showed a higher correlation.

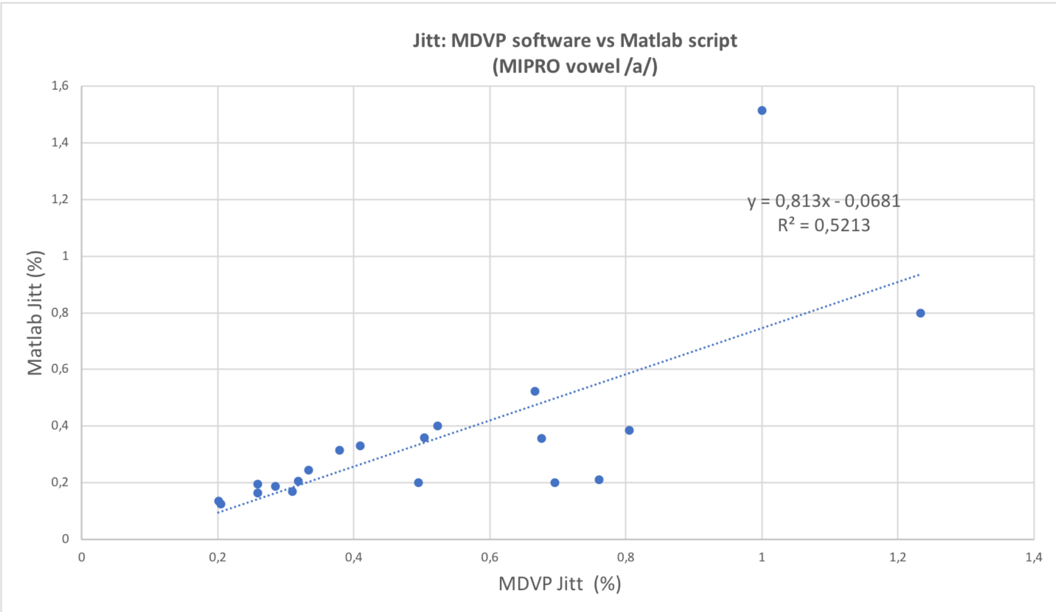

Figure 3.8: Dispersion Graphic with linear regression line of Jitt MDVP values and Jitt Matlab values (MIPRO vowel /a/).

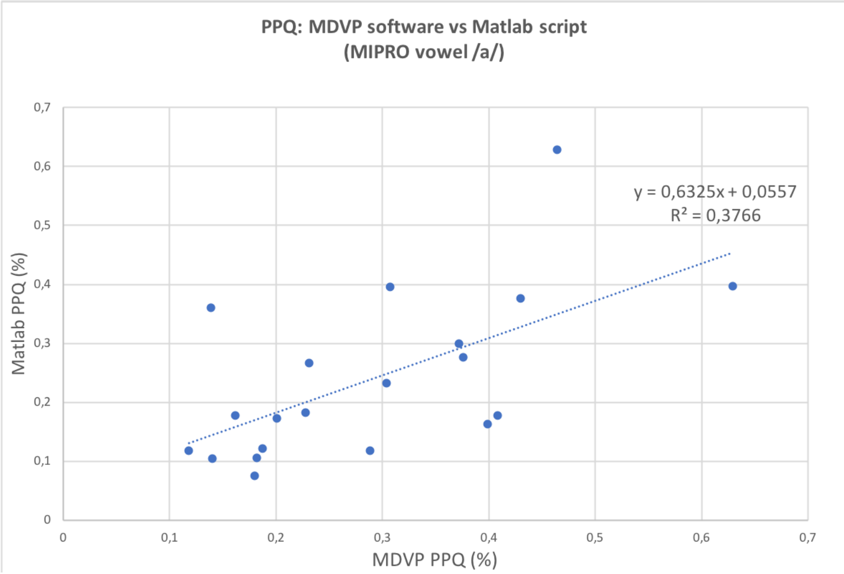

Figure 3.9: Dispersion Graphic with linear regression line of PPQ MDVP values and PPQ Matlab values (MIPRO vowel /a/).

Consequently to the results obtained, through the linear regression for each parameter the corresponding Matlab normality threshold was extracted, starting from the normality thresholds provided by MDVP software. In tab. 3.2 the thresholds are reported.

Parameter	***MDVP Threshold***	***Matlab Threshold***
Jitt	1,04	0,78
PPQ	0,84	0,59
Shim	3,81	3,63
vAm	8,20	8,16

Table 3.2: Summary table of normality threshold provided by MDVP software and the corresponding ones obtained in Matlab for MIPRO vowel /a/

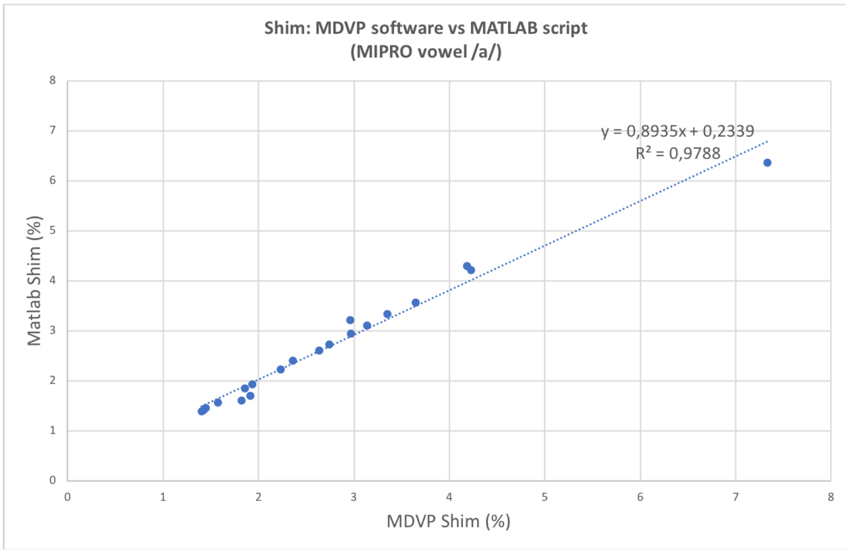

Figure 3.10: Dispersion Graphic with linear regression line of Shim MDVP values and Shim Matlab values (MIPRO vowel /a/).

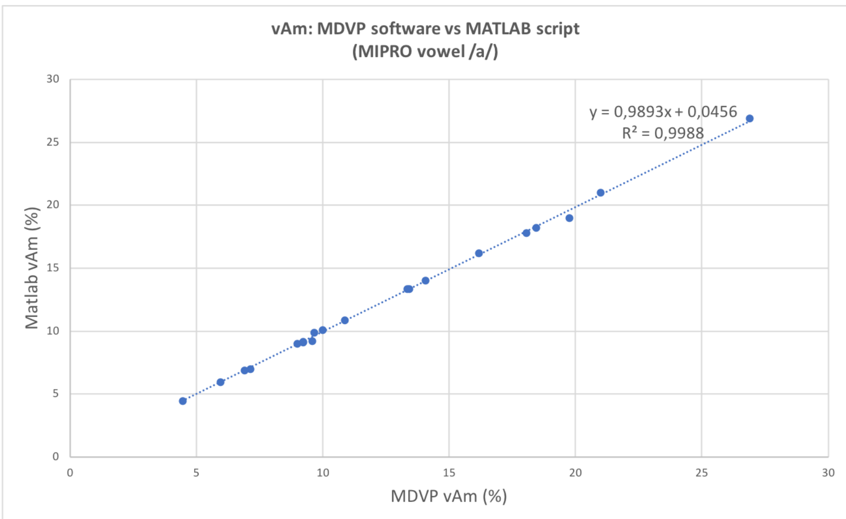

Figure 3.11: Dispersion Graphic with linear regression line of vAm MDVP values and vAm Matlab values (MIPRO vowel /a/).

3.6.3 Microphone in air: pre-surgery and post-surgery comparison

After the choice of the parameters subset and the corresponding threshold, the results obtained before and after surgery were compared, through different histograms. Each of them is associated to a parameter and for each subject three bars are plotted: the blue one represents the mean value of the parameters corresponding to the three vowels /a/ before the operation, and the orange and the green bars describes the corresponding mean value of parameters 1 month and 2 months after the operation respectively. On each bar the standard deviation bar is plotted. A low standard deviation indicates that the data points tend to be close to the mean of the three values of the parameter, while a high standard deviation indicates that the data are spread out over a wider range of values. Furthermore, in red are reported the thresholds.
In fig.3.12 it is described the Jitt (%) comparison. The *first subject*, the one who is affected by sulcus vocalis of the left vocal fold, has an anomalous behavior in terms of Jitt value, in fact before the operation the Jitt value is below the threshold, 1 month after the operation the parameter tends to decrease again, but after two months the value has an high rise compared to the pre-operation, above the threshold. This indicates that the subject has a less control of the frequency stability of the voice after the operation than before. The *second* and the *fifth subject*, affected respectively by Reinke's edema and Cyst of the left vocal fold, have very similar behavior in terms of Jitt: in the pre-surgery monitoring the parameter is very high (5 %) above the threshold but after one month and 2 months there is a noticeable lowering (0,5 %). The *third*, *fourth* and the *sixth subject*, the first two affected by polyp of the right vocal fold and last by cyst of the left vocal fold, in the pre-operation show very low Jitt (0,8 %), just above the threshold, and after the operation it further decreases around 0,2 %. Eventually, the *seventh* and the *eighth* subjects, both affected

in the pre-operation by Cysts and Edema of the left and the right vocal folds, shows very low and comparable jitt values in all three monitorings, so this parameter is not rilevant for them in discrimination between pre-surgery and post-surgery.

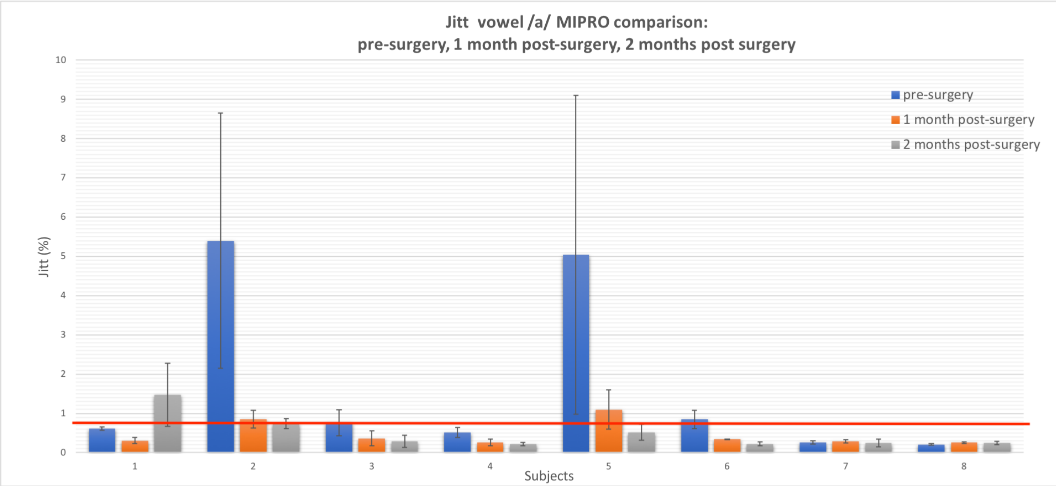

Figure 3.12: Jitt MIPRO comparison between pre-surgery, 1 month post-surgery and 2 months post-surgery vowels /a/ mean. In red is shown the threshold.

In fig.3.13 it is described the PPQ(%) comparison. The PPQ gives the variability from 5 to five periods of the fundamental period. The factor 5 of smoothing decreases the sensibility of the parameter to extraction errors of the fundamental period. The results are not so different from the Jitt ones. In particular, the *first subject*, has an anomalous behavior in terms of PPQ value, in fact 1 month after the operation the parameter tends to decrease, but after two months the value has an high rise compared to the pre-operation. The *second* and the *fifth subject*, have very similar behavior in terms of PPQ, but in this case the fifth subject has an higher PPQ value before the operation (7%) respect to the second one which is around 3,5%. The *sixth*, *fourth* show a very low decrease after 1 month and 2 months after the operation. The *third* subject show a rise of PPQ value after the operation but 2 months after the operation it decreases and it is lower than the pre-operation value. The *sev-*

enth subject is simillar to the third one but in the third monitoring the PPQ value settles on the value before the operation. Finally, the *eighth* show all the PPQ value around 1% in the three monitorings. Except for the first patient, in all cases the values are over the threshold, so can be seen an improvement in some subjects but the values are still too high to be considered healthy.

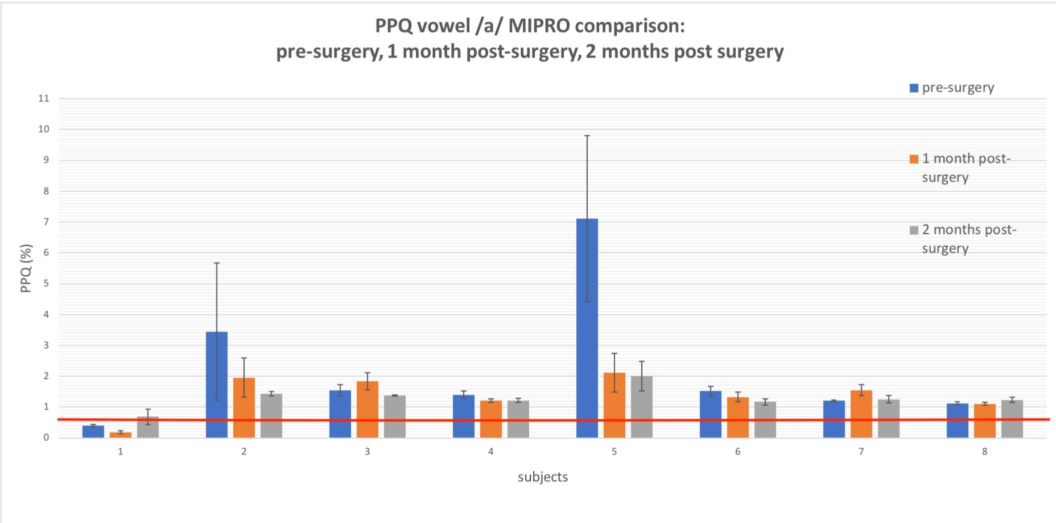

Figure 3.13: PPQ MIPRO comparison between pre-surgery, 1 month post-surgery and 2 months post-surgery vowels /a/. In red is shown the threshold.

In fig.3.14 it is described the Shim(%) comparison. The *first subject* has an anomalous behavior in terms of Shim value, in fact in the pre-operation monitoring the value is 5,8%, 1 month after the operation the parameter tends to decrease (3,5%) under the threshold, but after two months the value has an high rise compared to the pre-operation (7,5%). This indicates that the subject has a less control of the amplitude stability of the voice after the operation. The *second* and the *fifth* subjects have similar behavior in terms of Shim: the parameters have very high values in pre-surgery monitorings and there is a gradually decrease in the post-surgery monitorings, but all are always above the threshold. The *third*, *fourth* and *sixth* subjects show a Shim decrease after the operation, starting from a 5% of Shim in pre-surgery monitoring to shim values under the threshold. The *seventh* and the *eighth* show similar behaviours in all the monitorings.

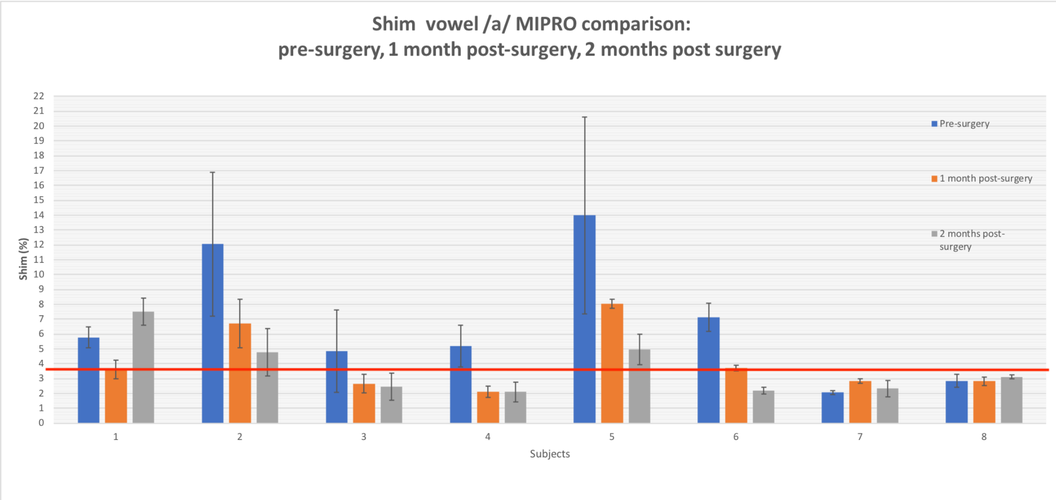

Figure 3.14: Shim MIPRO comparison between pre-surgery, 1 month post-surgery and 2 months post-surgery vowels /a/.In red is shown the threshold.

The fig. 3.15 shows the comparison pre and post-surgery of the vAm (%) parameter, the results do not give relevant information in discrimination of pre and post-surgery situation. Furthermore, all the parameter values are above the threshold both before and after the operation.

In fig. 3.16 is shown the HNR MIPRO comparison between pre-surgery, 1 month post-surgery and 2 months post-surgery vowels /a/. The *first* subject show an HNR value raise 1 month after the operation but 2 months after the operation it decreases again. From the *second* to the *sixth* patient there is a reise of HNR, as it is expected the harmonics component increases and the noise component decreases after the operation. Instead, the *seventh* and the *eighth* show a decrease of HNR.

In fig. 3.17 are reported the CPPS distribution of vowel /a/ of the first and the second patient. For all patients three graphs are reported because they represents the three repetitions of the vowel /a/ and in each repetition the

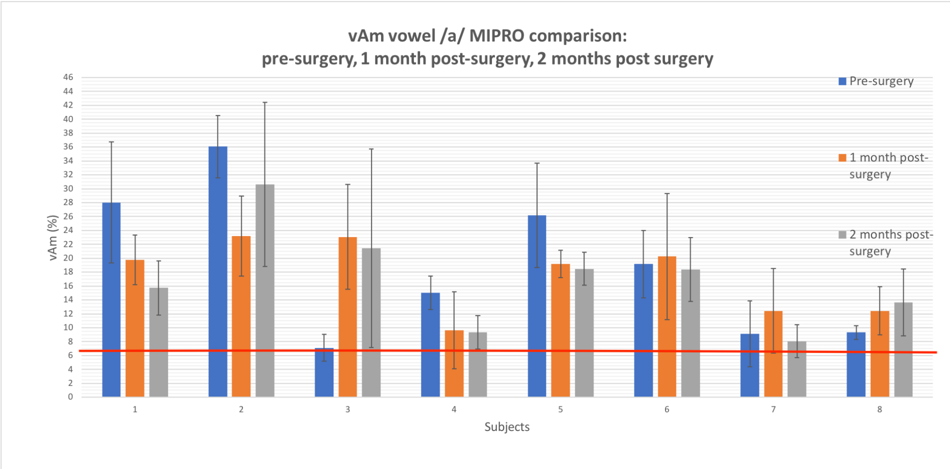

Figure 3.15: vAm MIPRO comparison between pre-surgery, 1 month post-surgery and 2 months post-surgery vowels /a/.

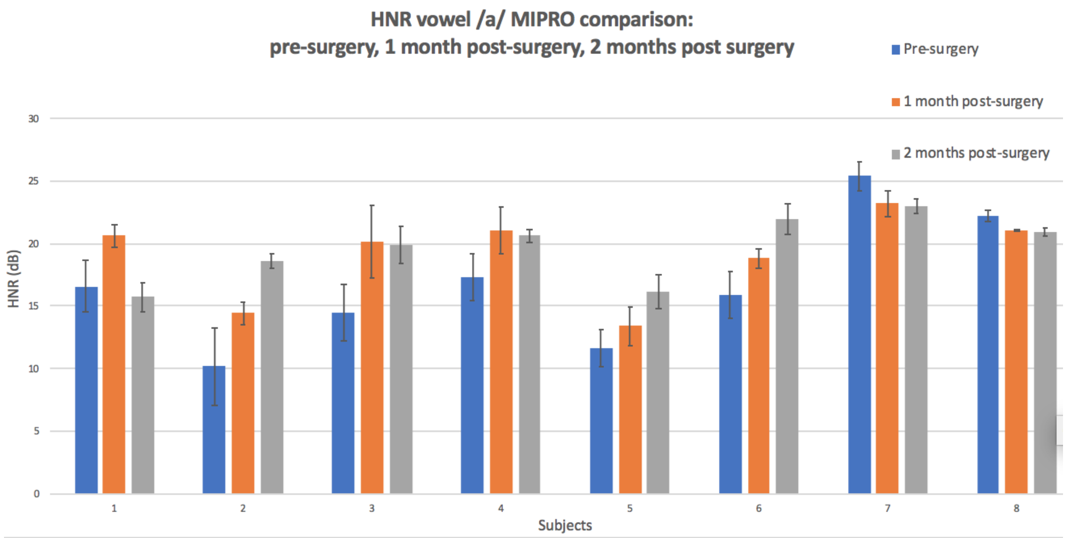

Figure 3.16: HNR MIPRO comparison between pre-surgery, 1 month post-surgery and 2 months post-surgery vowels /a/.

pre-surgery (in blue), 1 month post-surgery (in orange) and 2 months post-surgery (in yellow) distribution are overlapped, in order to see the distribution changes. The first /a/ of the fist patient looks to rise 1 month after operation, from 14 dB of $CPPS_{\mathrm{mean}}$ to 18 dB but 2 months after the operation it decrease again. The other two /a/ look similar in all three monitoring with a CPPSmean around 15dB. For the second patient, in the first monitoring the CPPS distribution looks bimodal, tipical of dysphonia, with some CPPS low value between 4 dB and 10 dB. After the rehabilitation, the CPPS occurrences moved to higher dB values and it is more remarkable two months after surgery, showing an healthy condition. The results are very similar in the three /a/ repetition. In fig. 3.18 the third patient have CPPS distribution occurrences that decrease 2 month after operation with $CPPS_{\text{mean}}$ around 15 dB. The CPPS distribution occurrences belonging to the fourth patient raise after the operation with values close to 18 dB in all three sustained vowels. The same happens for the fifth and sixth patient (fig. 3.19), with a post operation distribution (in yellow) that moves to higher dB values. Eventually, the seventh patient shows a decrease in the CPPS distribution in all three sustained vowels /a/ and the eighth patient shows distributions very similar to each other in all the monitorings (fig. 3.20).

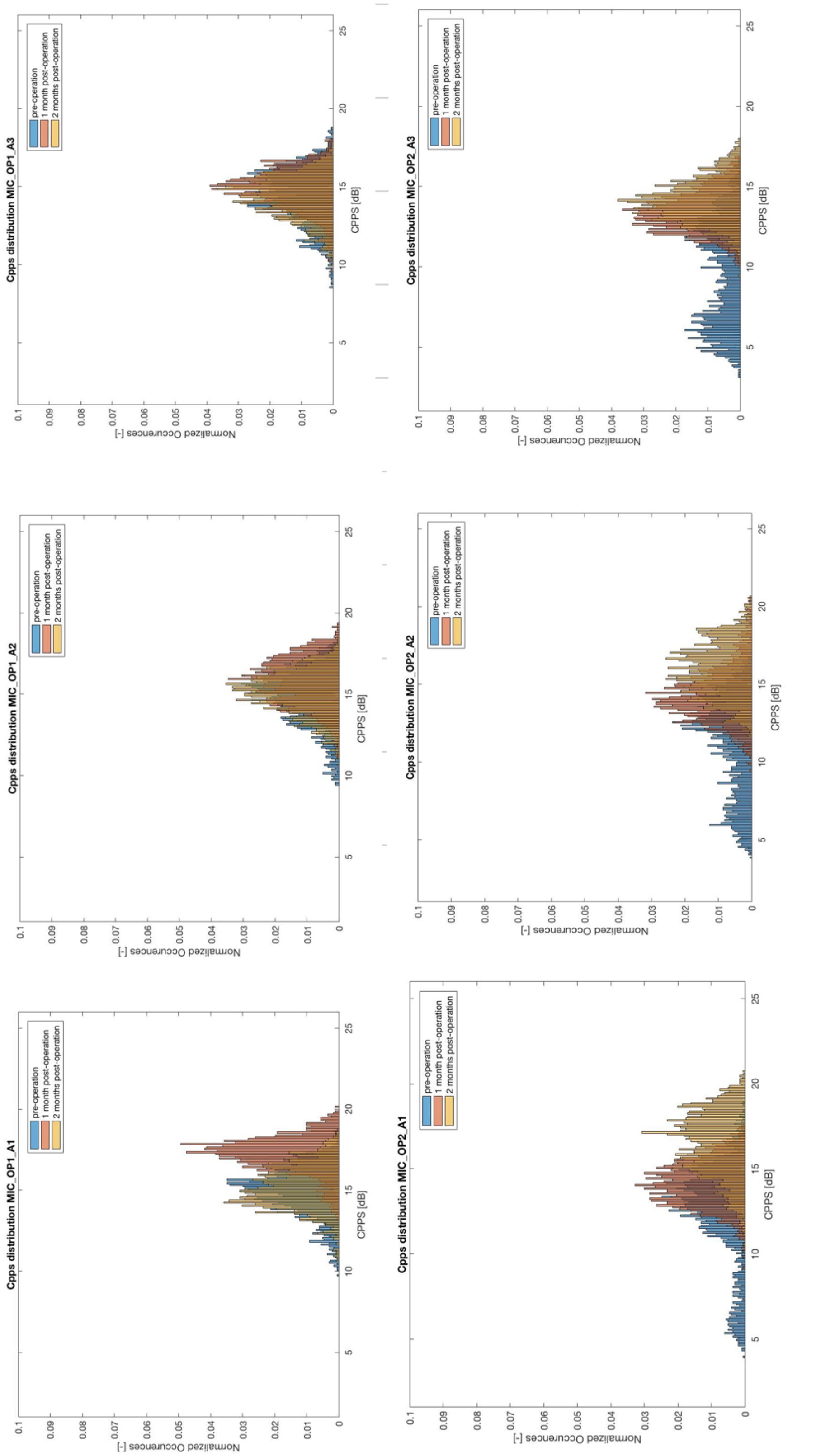

Figure 3.17: CPPS MIPRO distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery, 1 month post surgery and 2 months post surgery. The three graphs above the image belongs to the first patient, the three down belongs to the second patient.

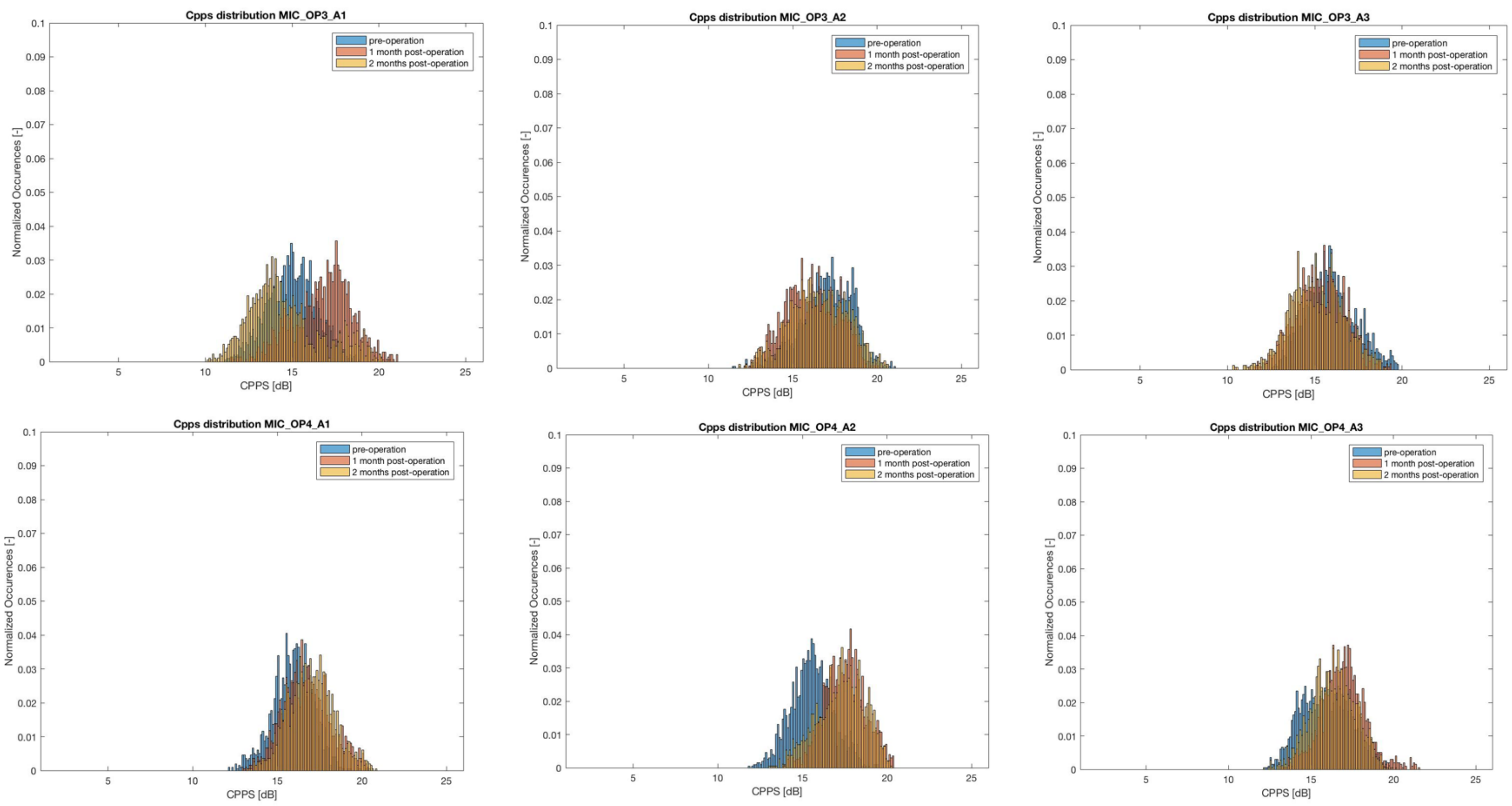

Figure 3.18: CPPS MIPRO distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery, 1 month post surgery and 2 months post surgery. The three graphs above the image belongs to the third patient, the three down belongs to the fourth patient.

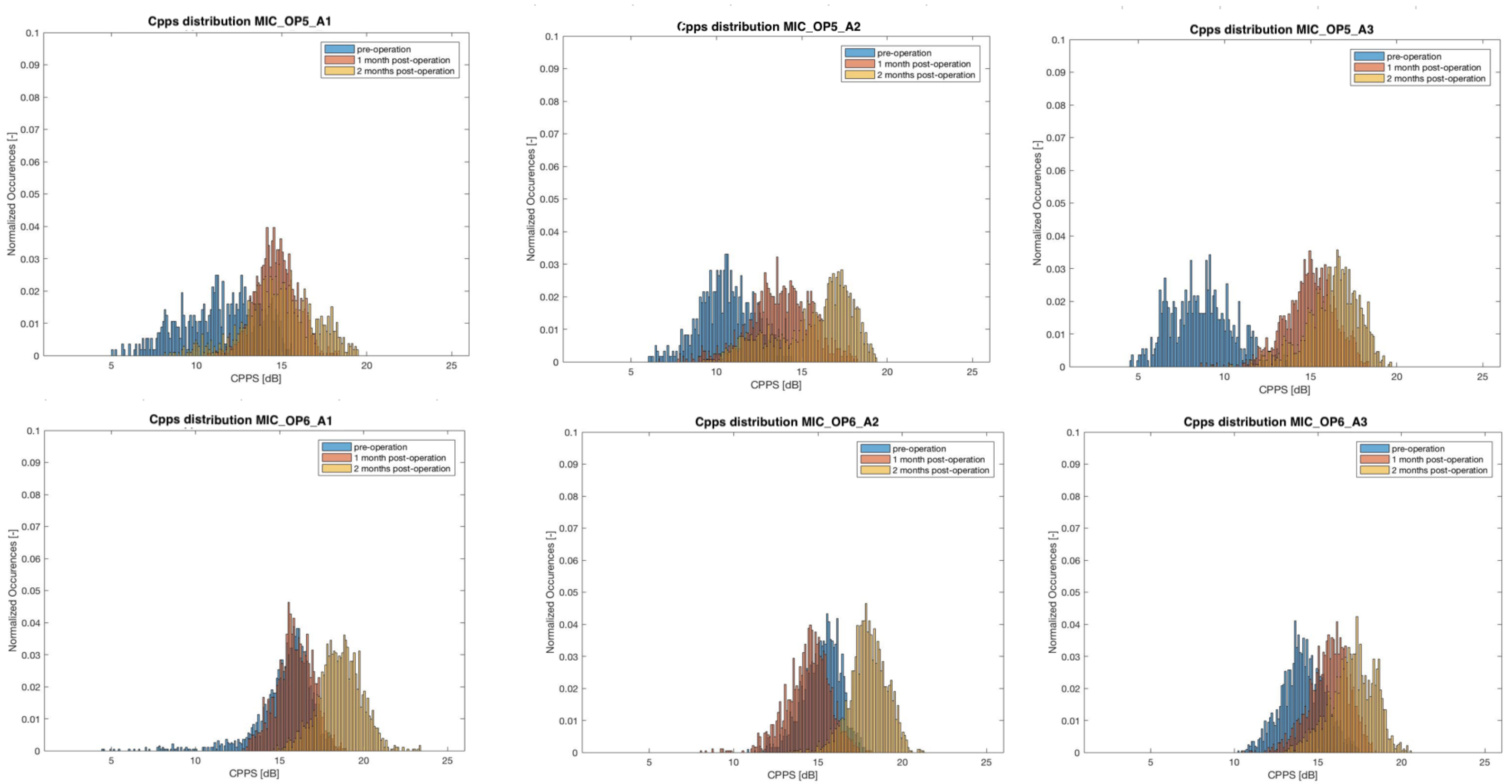

Figure 3.19: CPPS MIPRO distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery, 1 month post surgery and 2 months post surgery. The three graphs above the image belongs to the fifth patient, the three down belongs to the sixth patient.

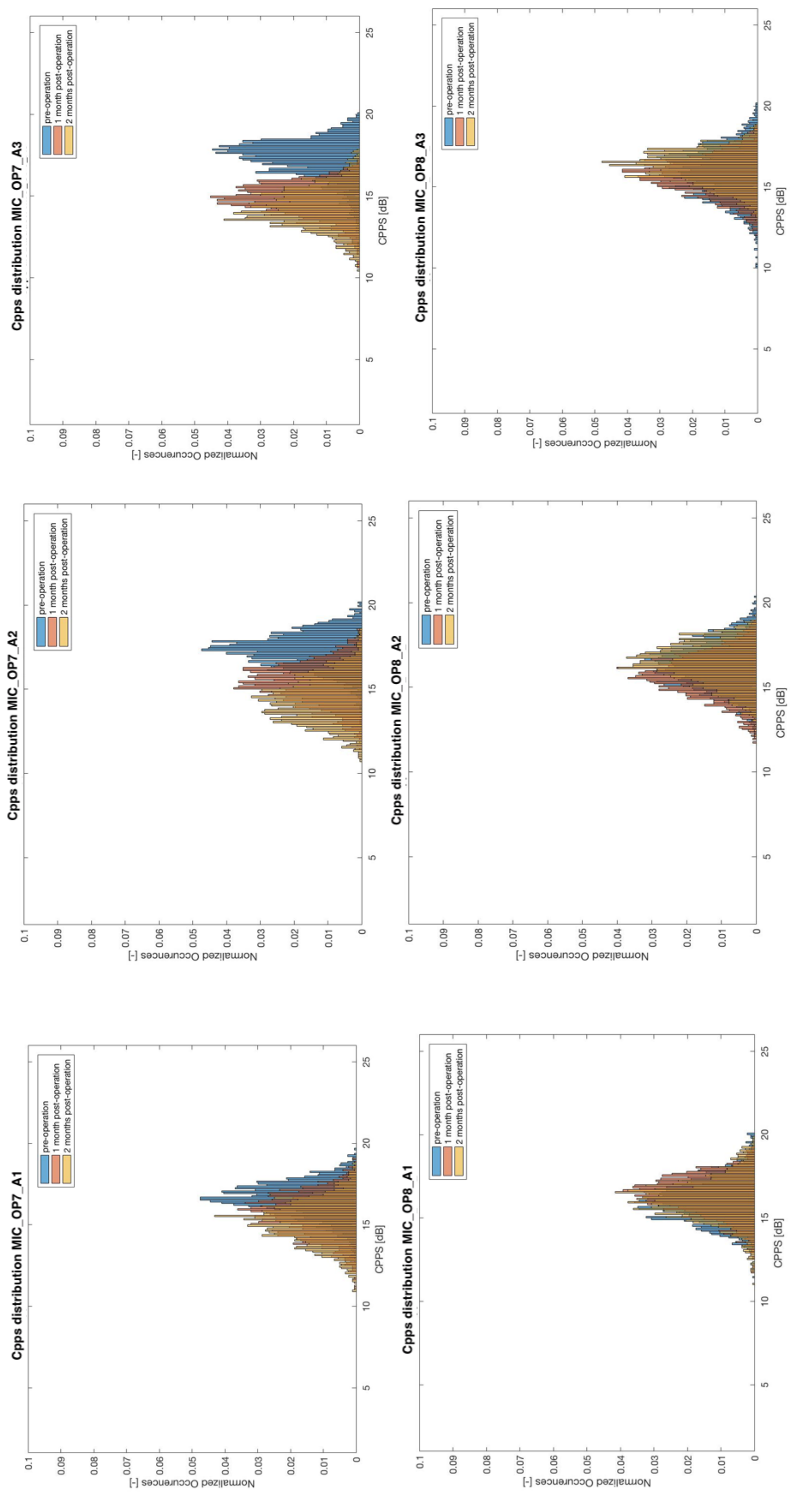

Figure 3.20: CPPS MIPRO distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery, 1 month post surgery and 2 months post surgery. The three graphs above the image belongs to the seventh patient, the three down belongs to the eighth patient.

In figure 3.21 is reported the CPPS 5prc MIPRO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean. In red is showed the threshold, equals to 15 dB, for discrimination between healthy and unhealthy provided by Castellana et al.[14]. It is a localization parameter, so the graph shows how the CPPS distribution moves after the operation to lower or higher dB values. The first, second, fourth, fifth, sixth, and eighth patients show a rise of CPPS 5prc after the operation, of which significant differences are noted above all in second, fifth and sixth patients. Instead, the third and the seventh patients show a decrease of the parameter after the operation. Only for the sixth patient the value of the parameter is above the threshold provided. In fig. 3.22 is reported

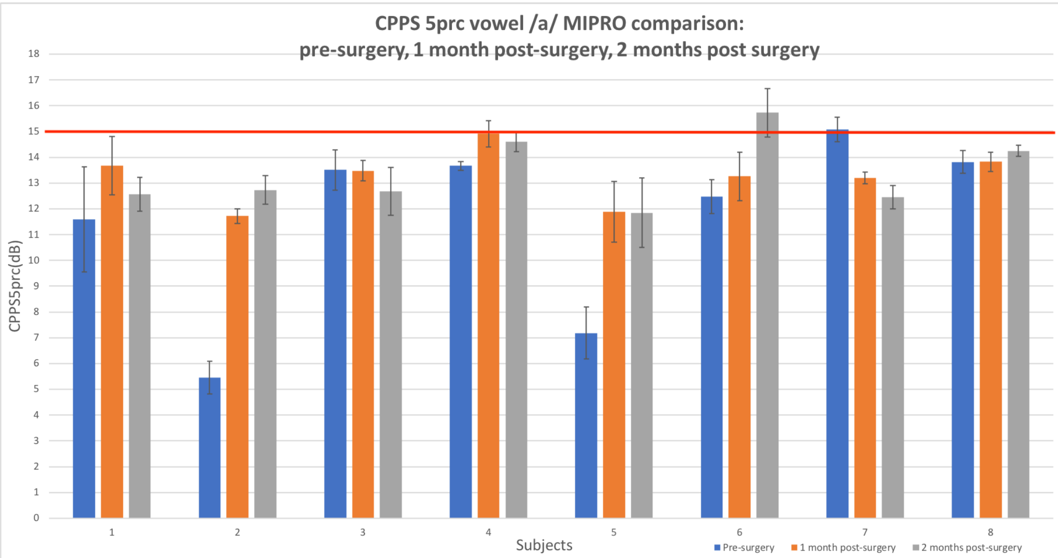

Figure 3.21: CPPS 5prc MIPRO comparison between pre-surgery, 1 month post-surgery and 2 months post-surgery vowels /a/ mean. The threshold is showed in red.

the $CPPS_{\text{std}}$ comparison. A decrease of the $CPPS_{\text{std}}$ parameter is a symptom of post-operation improvement. The higher discrimination is seen for the second patient from a value of 3 dB before the operation to a value of 1,5 dB after the operation. For the other patients the results are similar, not so relevant, in all the situations.

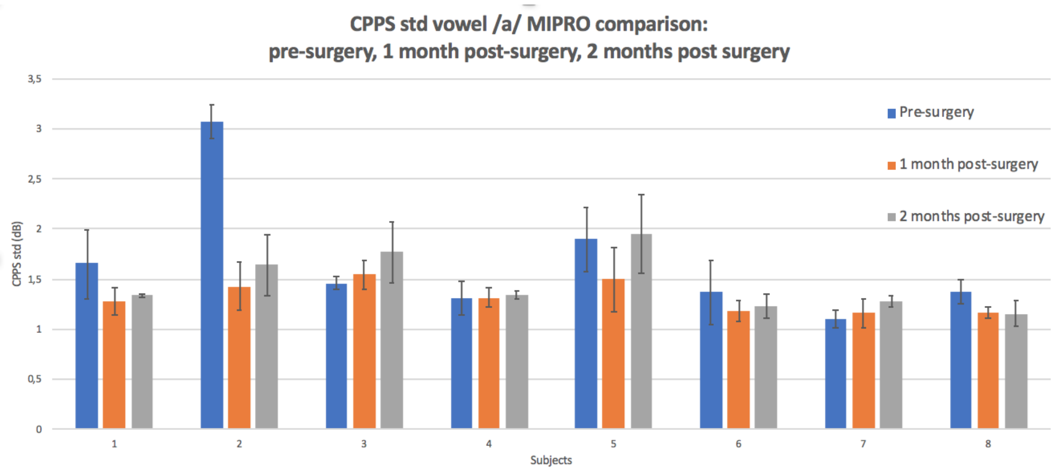

Figure 3.22: CPPSstd MIPRO comparison between pre-surgery, 1 month post-surgery and 2 months post-surgery vowels /a/ mean.

3.6.4 Piezoelectric contact microphone: pre-surgery and post-surgery comparison

After the choice of the parameters subset, the results obtained by piezoelectric microphone before and after surgery were compared, through different histograms. Each of them is associated to a parameter and for each subject three bars are plotted: the blue one represents the mean value of the parameters corresponding to the three vowels /a/ before the operation, and the orange and the green bars describes the corresponding mean value of parameters 1 month and 2 months after the operation respectively. On each bar the standard deviation bar is plotted. The threshold provided for MIPRO microphone were used also for the PIEZO microphone results in vowel /a/.
In figure 3.23 it is described the Jitt (%) comparison. The *first* subject, the one who is affected by sulcus vocalis of the left vocal fold, has an anomalous behavior in terms of Jitt value, in fact 1 month after the operation the param-

eter tends to decrease, but after two months the value has an high rise above the threshold, also compared to the pre-operation (the same happens for the MIPRO microphone). This indicates that the subject has a less control of the frequency stability of the voice after the operation. The *second* subject shows a jitt value of 2,6 % in the pre-operation phase. The parameter has an high rise 1 month after the operation (5,8%) but after 2 months it decrease again (1,5%). The *third*, the *fifth*, the *sixth* subjects have similar behavior in terms of Jitt: in the pre-surgery monitoring the parameter is higher than the second and the third monitoring after the operation, in particular the fifth subject shows an higher jitt value of 4 % that in the post-operation goes below the threshold. The *fourth* in the pre-operation show very low jitt (0,2%) and after the operation it further decreases at 0,1 %. Eventually, the *seventh* and the *eighth* subjects, both affected in the pre-operation by Cysts and Edema of the left and the right vocal folds, shows very low and comparable Jitt values below the threshold in all three monitorings.

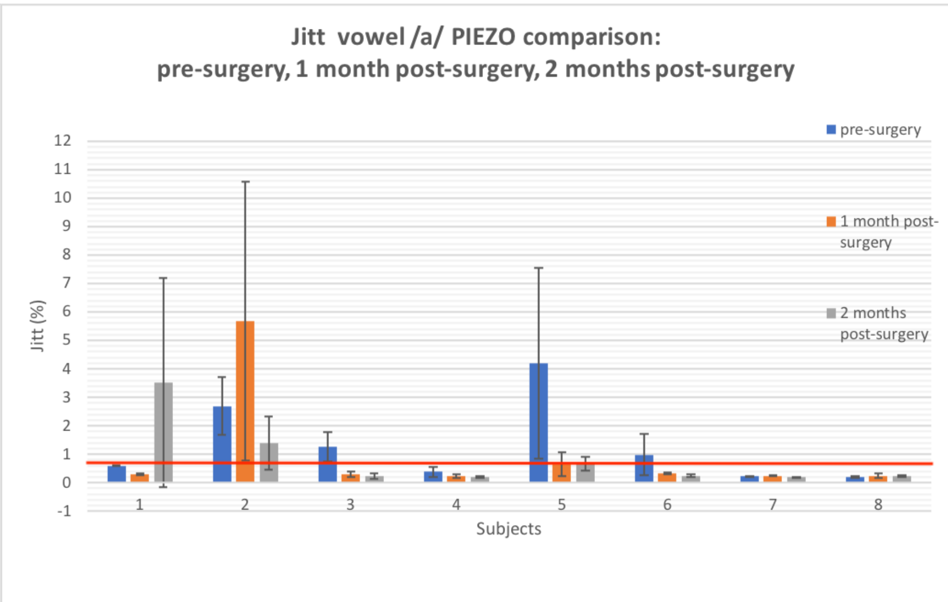

Figure 3.23: Jitt PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean.

In fig.3.24 it is described the PPQ(%) comparison. The PPQ gives the variability from 5 to five periods of the fundamental period. The factor 5 of smoothing decreases the sensibility of the parameter to extraction errors of the fundamental period. The *first* subject, has an anomalous behavior in terms of PPQ value, in fact 1 month after the operation the parameter tends to decrease, but after two months the value has an high rise compared to the pre-operation. Also the *second* has a raise in terms of PPQ after the operation and it decrease after 2 months but it is always higher than the pre-operation. The *third*, the *sixth* and the *seventh* subject, have very similar behavior in terms of PPQ: in the pre-operation the value is close to 3% but 1 month after the operation it raise and after 2 months it decreases around the pre-operation value. The *fifth* subject shows a very high PPQ value in the pre-operation around 11% and it decrease in the next monitorings around 4 %. Eventually, the *fourth* and the *eighth* show a gradually decrease of PPQ value in the post-operation monitorings. In all cases the parameter values are over the threshold.

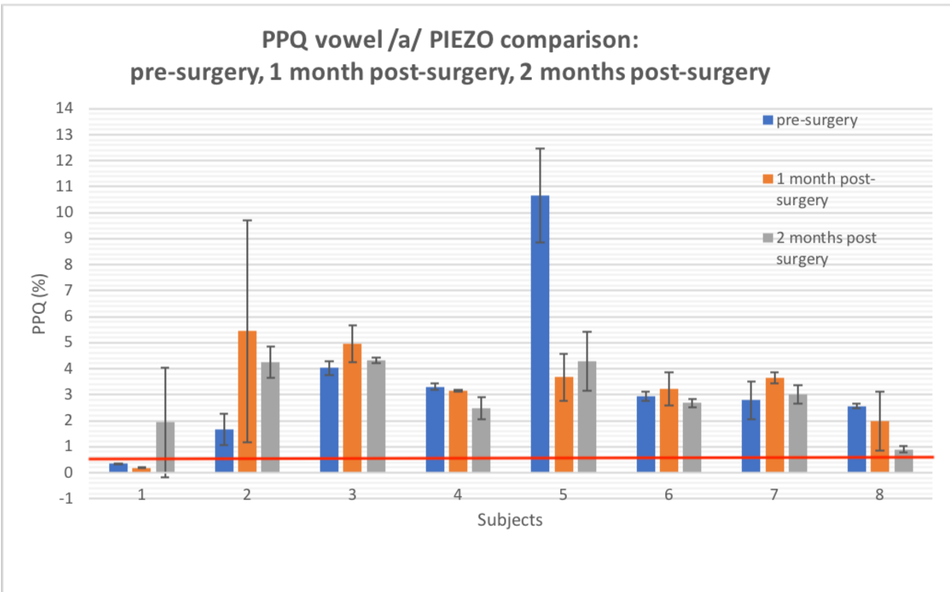

Figure 3.24: PPQ PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean and the normality threshold is showed in red.

In fig.3.25 it is described the Shim(%) comparison. The *first* subject has an anomalous behavior in terms of Shim value, in fact in the pre-operation monitoring the value is 3% below the threshold, 1 month after the operation the parameter decreases at 2%, but after two months the value has an high rise compared to the pre-operation 11%. This indicates that the subject has a less control of the amplitude stability of the voice after the operation. The *second* and the *fifth* subjects have similar behavior in terms of Shim: the parameters show very high values in pre-surgery monitorings and there is a gradually decrease in the post-surgery monitorings. The *third*, *fourth* and *sixth* subjects show a Shim decrease after the operation that increase after 2 months. The *seventh* shows similar behaviours in all the monitorings. The *eighth* subject shows a rise in terms of Shim after the operation at 2,5% that decrease again 2 months after, equaling the pre-operation. The best discrimination between pre and post operation is given for the second and the fifth patients but the values before the operation are steel above the threshold, which means they are still considered unhealthy.

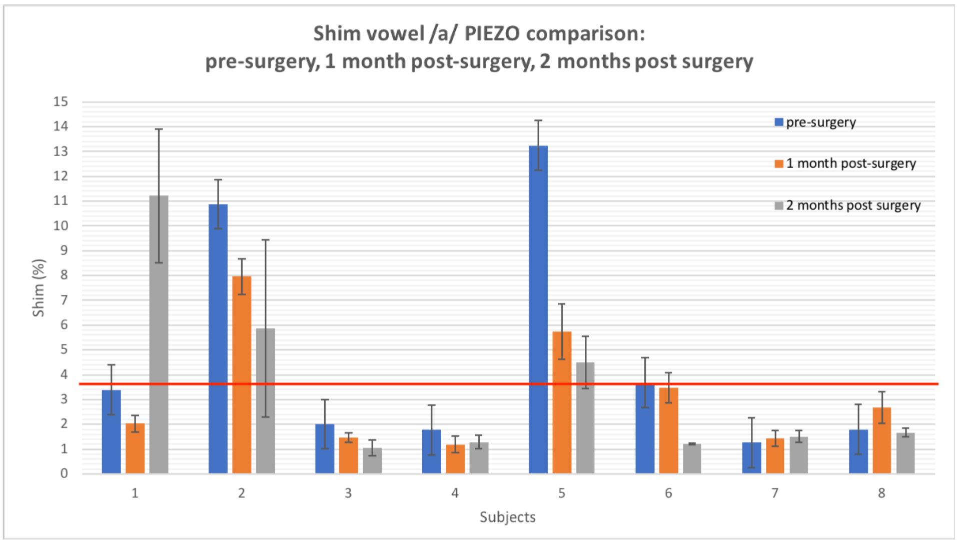

Figure 3.25: Shim PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean. In red is showed the threshold.

The vAm parameter (fig.3.26) gives the pick amplitude variation. It increases for any variation in amplitude, whether random or regular, whether short or long term. For the *first* patient the vAm behaviour is similar to Shim one with an increase of the parameter after two months post-surgery. The *second* patient's vAm shows a decrease after 1 month and an increase after 2 months but lower than the pre-operation. The *third* and the *sixth* patients shows a increase of the parameter after the operation. The *fourth* patient's vAm shows very similar values in the three monitorings and the *seventh* and the *eighth* patients show a gradually increase of the parameter after the operation. Except for the second and fifth patient, the parameter do not give relevant information in discriminating between pre and post-operation, also considering the threshold.

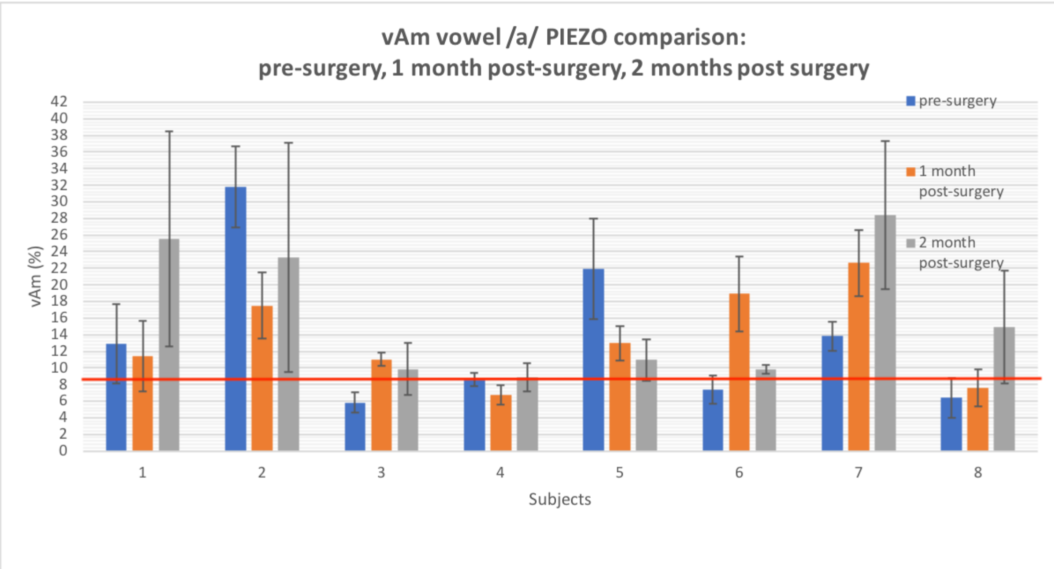

Figure 3.26: vAm PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean.

The HNR parameter (fig. 3.27) shows for the *second*, *third*, *fifth*, *sixth* and *eighth* subjects an increase of harmonic components and a decrease of noise component 2 months after the operation. The *fourth* and the *seventh* show very high values in all the three monitoring around 27 dB. Only the *first* subject shows a decrese from 23 dB to 15 dB after the operation.

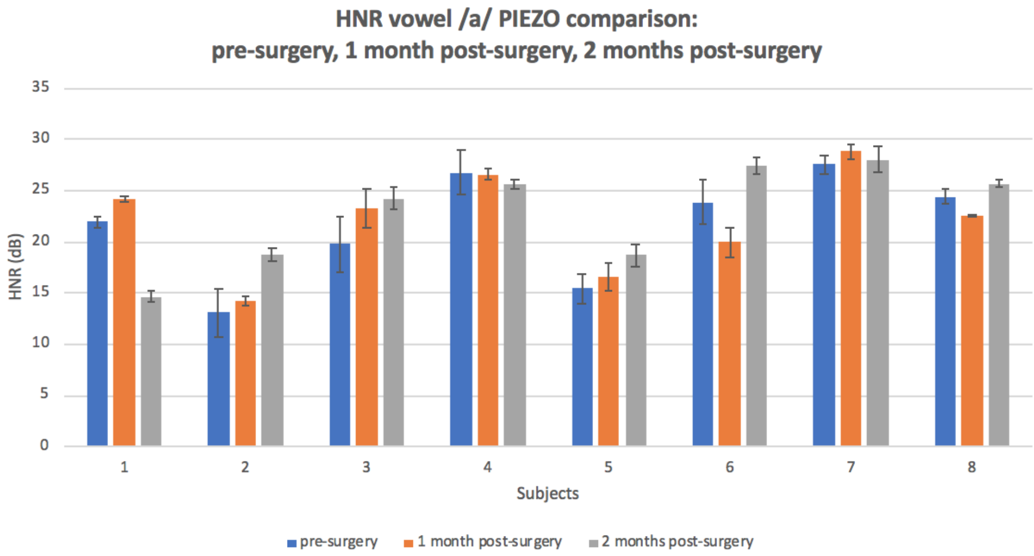

Figure 3.27: HNR PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean.

In fig.3.28 are reported the CPPS PIEZO distribution of vowel /a/ of the first and the second patient. For all patients three graphs are reported because they represent the three repetitions of the vowel /a/ and in each repetition the pre-surgery (in blue), 1 month post-surgery (in orange) and 2 months post-surgery (in yellow) distribution are overlapped, in order to see the distribution changes. The first /a/ of the fist patient looks to rise 1 month after operation, from 16 dB of CPPSmean to 18 dB but 2 months after the operation it decrease at 13 dB. The other two vowels /a/ comparison look similar to the first one. For the second patient, in the first monitoring the CPPS distribution looks bimodal, typical of dysphonia, with some CPPS low value between 4 dB and 10 dB. After the rehabilitation, the CPPS occurrences moved to higher dB values and it is more remarkable two months after surgery. The results are very similar in the three /a/ repetition and to the MIPRO vowel /a/ results. In fig. 3.29 the third patient have CPPS distribution occurrences that increase 2 months after operation with CPPSmean around 18 dB. The CPPS distribution occurrences belonging to the fourth patient raise 1 month after the operation around 19 dB in all three sustained vowels, but after 2 months it decreases again with a value of 17 dB. The fifth and sixth patients (fig. 3.30) show a post operation distribution (in yellow) that moves to higher dB values. Eventually, the seventh patient shows a similar CPPS distribution before and 2 months after the operation in all three sustained vowels /a/ and the eighth patient shows distributions with lower CPPSmean values after the operation in all the monitorings (fig. 3.31).

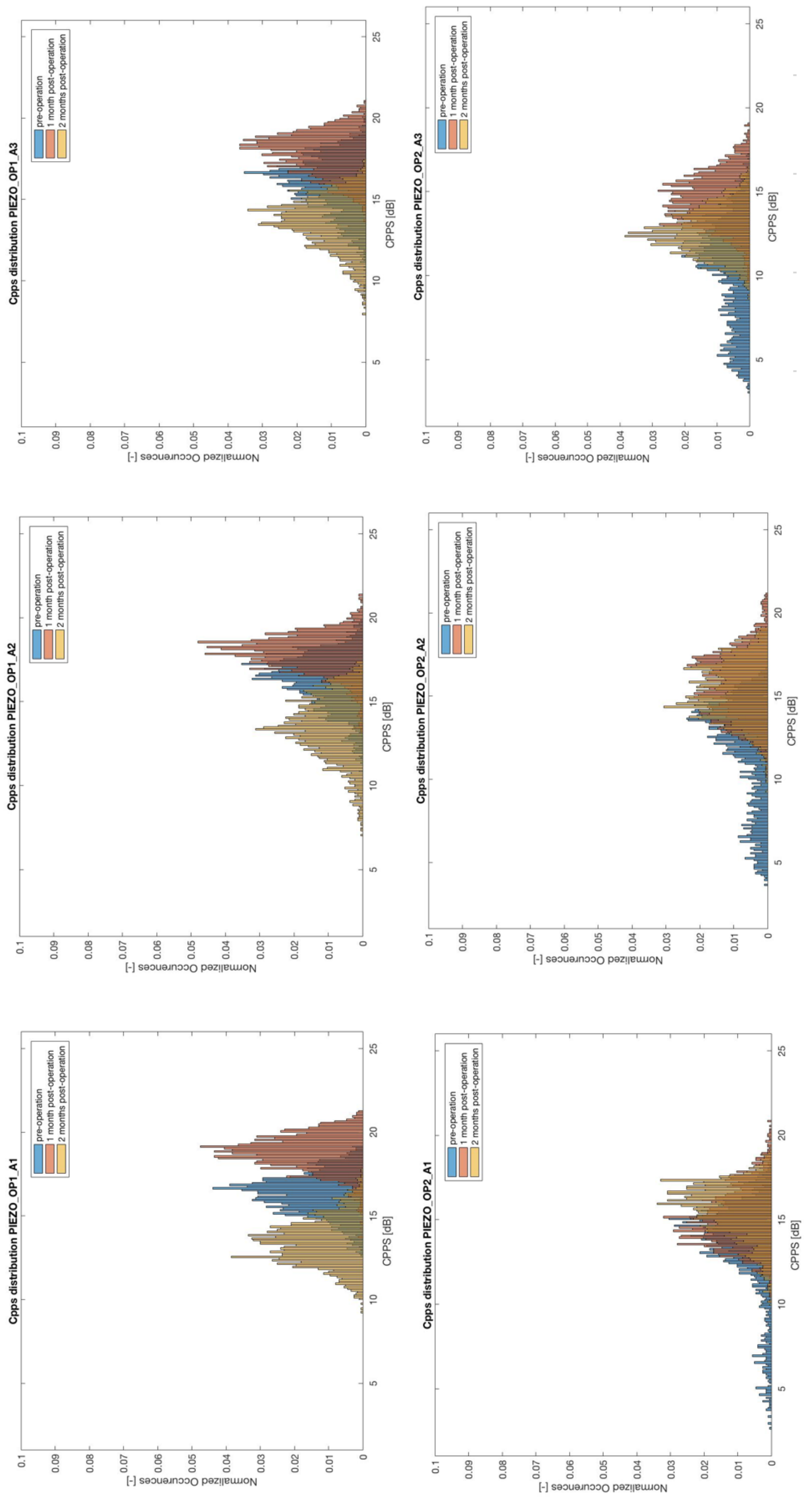

Figure 3.28: CPPS piezo distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery (blue), 1 month post surgery (orange) and 2 months post surgery (yellow). The three graphs above the image belongs to the first patient, the three down belongs to the second patient.

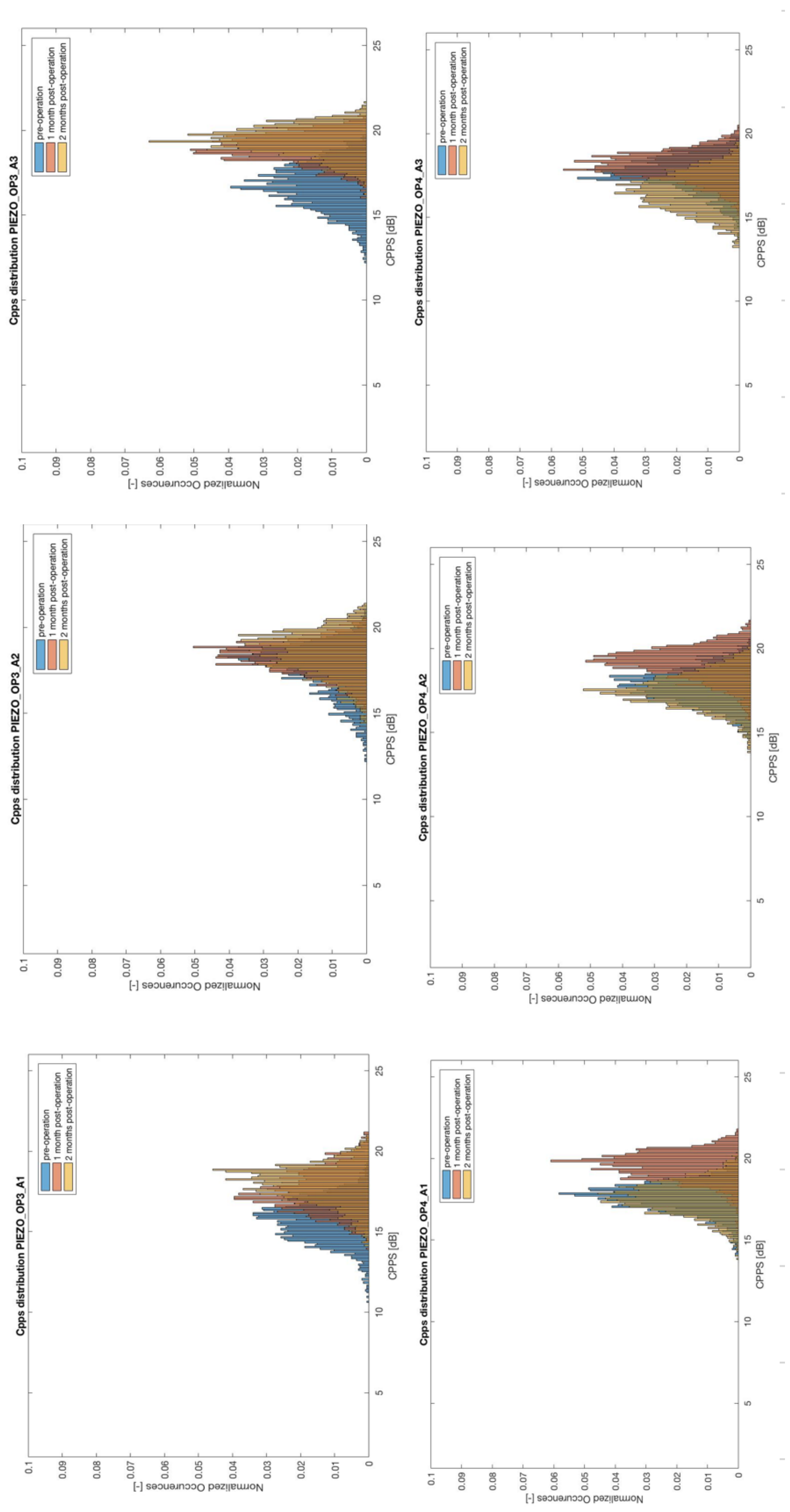

Figure 3.29: CPPS piezo distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery (blue), 1 month post surgery (orange) and 2 months post surgery (yellow). The three graphs above the image belongs to the third patient, the three down belongs to the fourth patient.

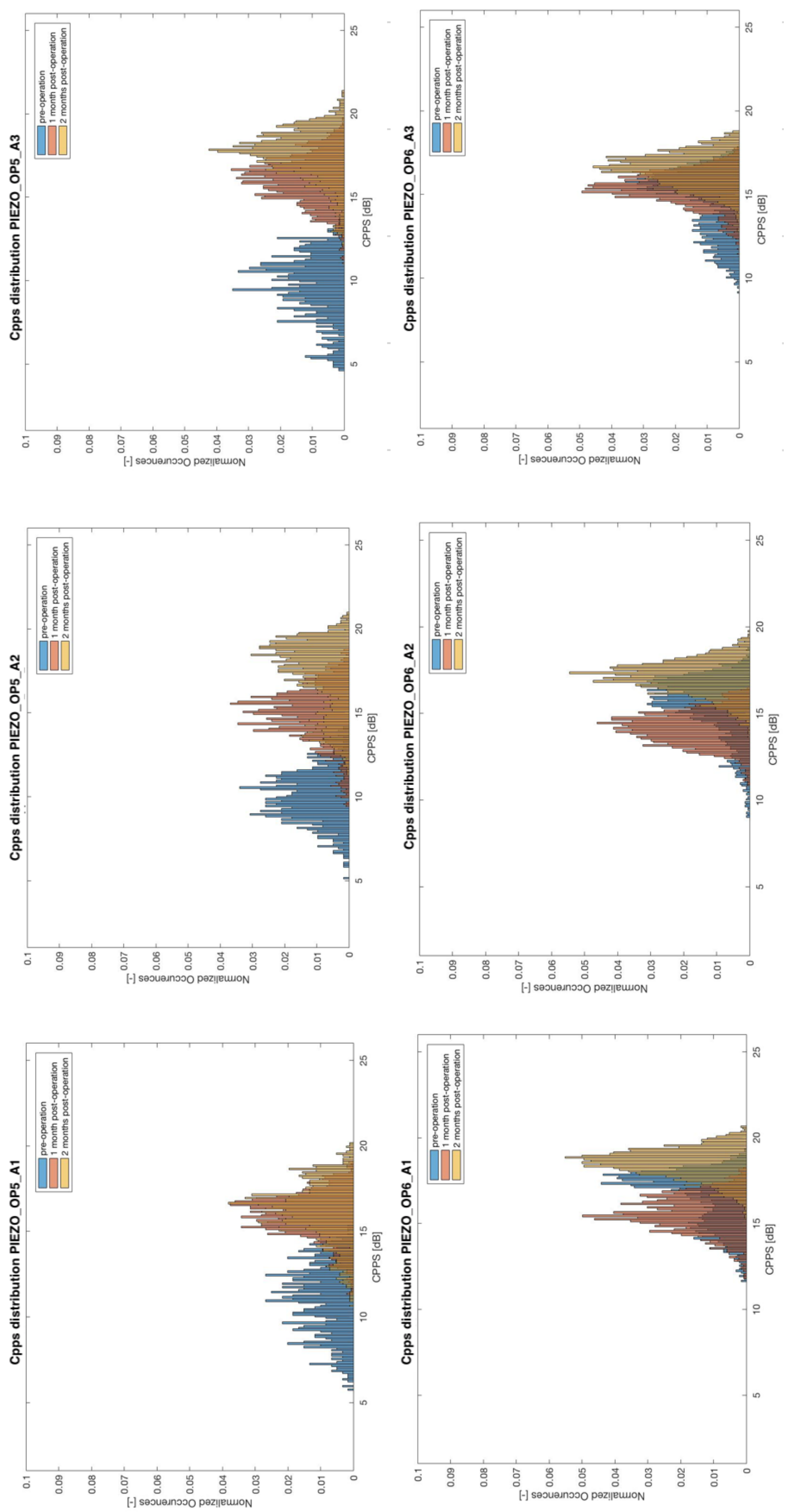

Figure 3.30: CPPS piezo distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery (blue), 1 month post surgery (orange) and 2 months post surgery (yellow). The three graphs above the image belongs to the fifth patient, the three down belongs to the sixth patient.

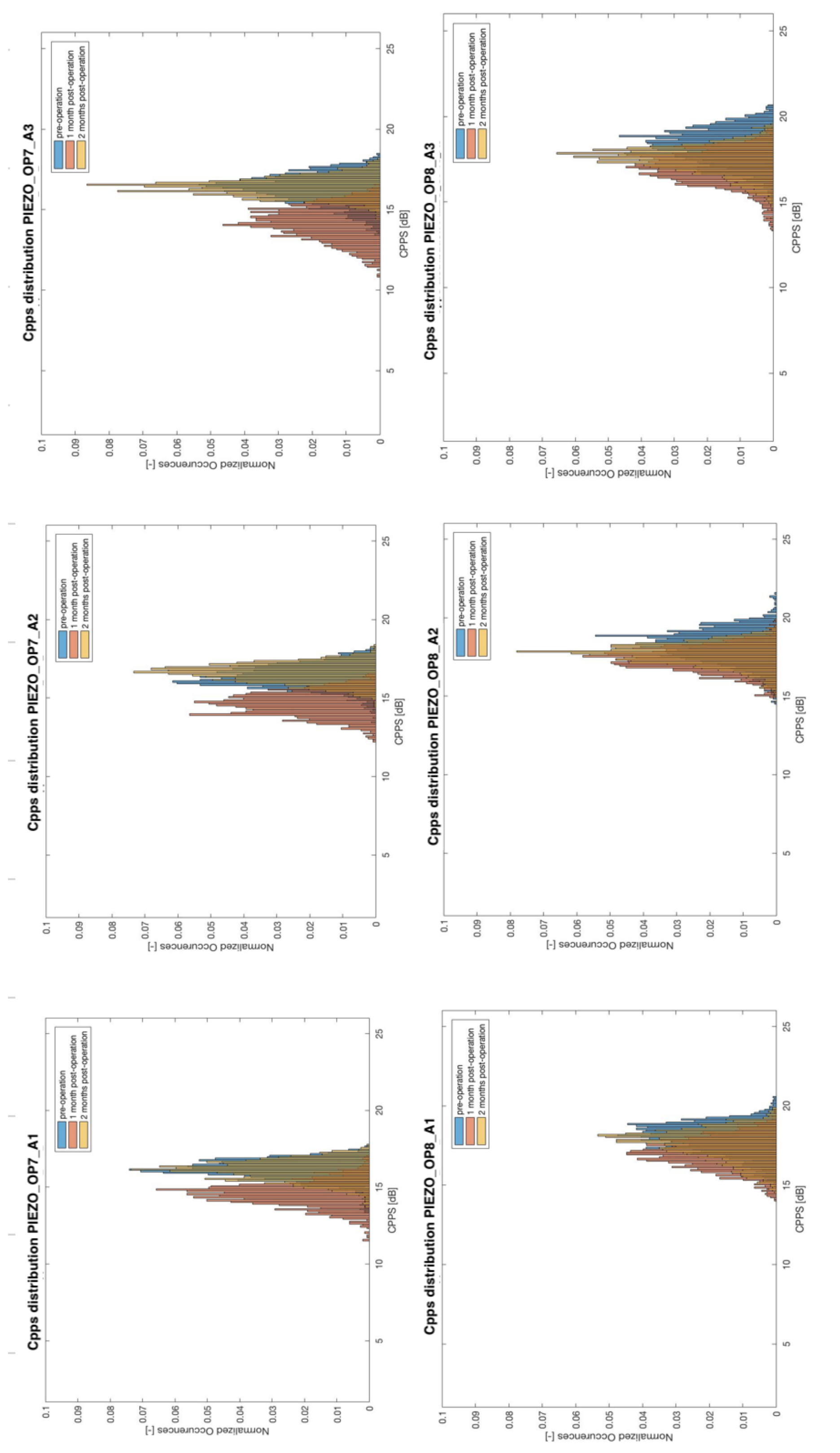

Figure 3.31: CPPS piezo distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery (blue), 1 month post surgery (orange) and 2 months post surgery (yellow). The three graphs above the image belongs to the seventh patient, the three down belongs to the eighth patient.

In fig.(3.32) is reported the CPPS 5prc PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/, with the threshold equals to 14,7 [13]. It is a localization parameter, so the graph shows how the CPPS distribution moves after the operation to lower or higher dB values. In particular the second, the fifth and the sixth patients show relevant CPPS 5prc improvement after operation. But in the first two cases the values are always under the threshold. In fig.3.33

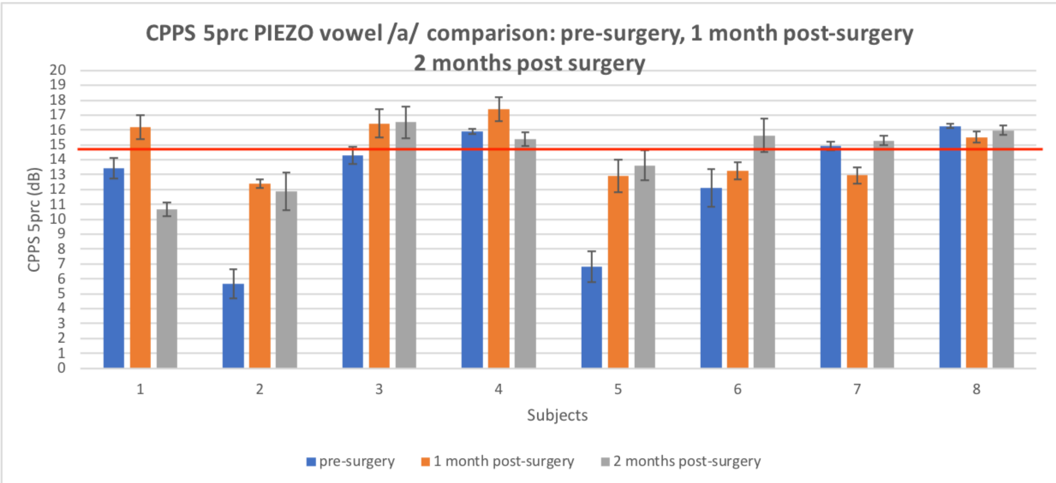

Figure 3.32: CPPS5th PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean. In red is shown the threshold.

are shown the CPPSstd PIEZO results. The graph shows similar results to the MIPRO ones. In particular relevant lowering of the parameter are obtained only for the second patient.

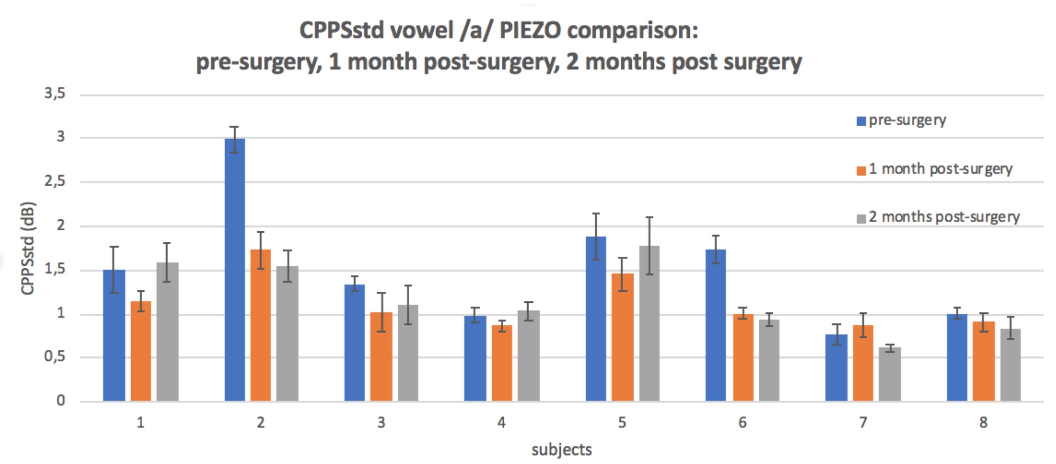

Figure 3.33: CPPSstd PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean.

3.6.5 Electret Condenser Microphone: pre-surgery and post-surgery comparison

After the choice of the parameters subset, the results obtained by Electret Condenser Microphone (ECM) before and after surgery were compared, through different histograms. The ECM recordings were partly lost due to a malfunction of the recorder, so only four of the eight subjects involved in the study were submitted to the ECM analysis and they were: the third, the fourth, the sixth and the seventh. The same thresholds given for the MIPRO vowel /a/ parameters are also used in the ECM results discrimination.

In fig. 3.34 is shown the Jitt comparison. Relevant results are obtained for

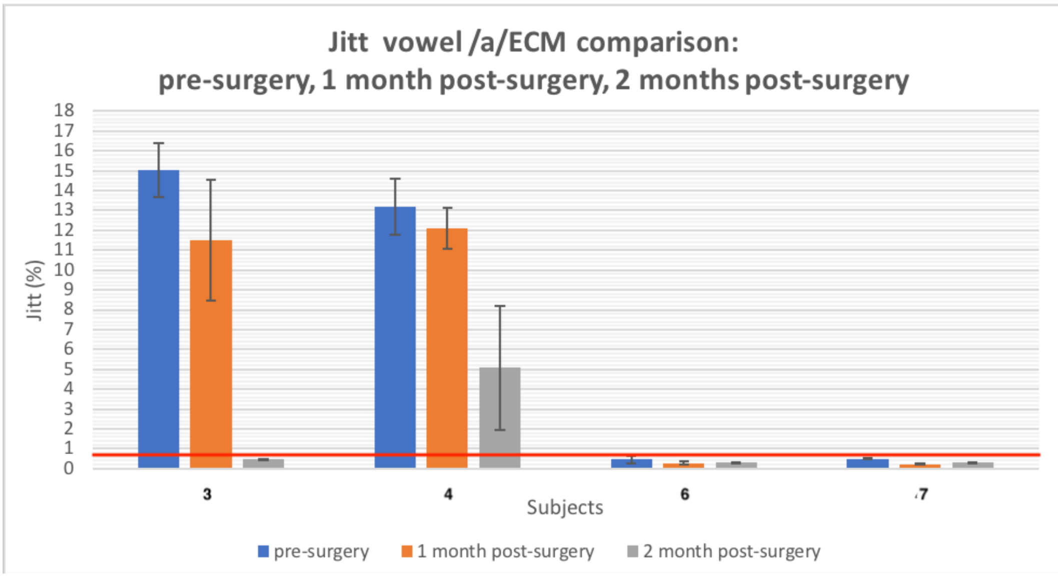

Figure 3.34: Jitt ECM comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean. In red is shown the threshold.

the third and the fourth patients but only the first with a post-operation Jitt value below the threshold. The sixth and seventh patients show very similar low values in all three monitorings.

Similar results are obtained for PPQ parameter (fig.3.35).

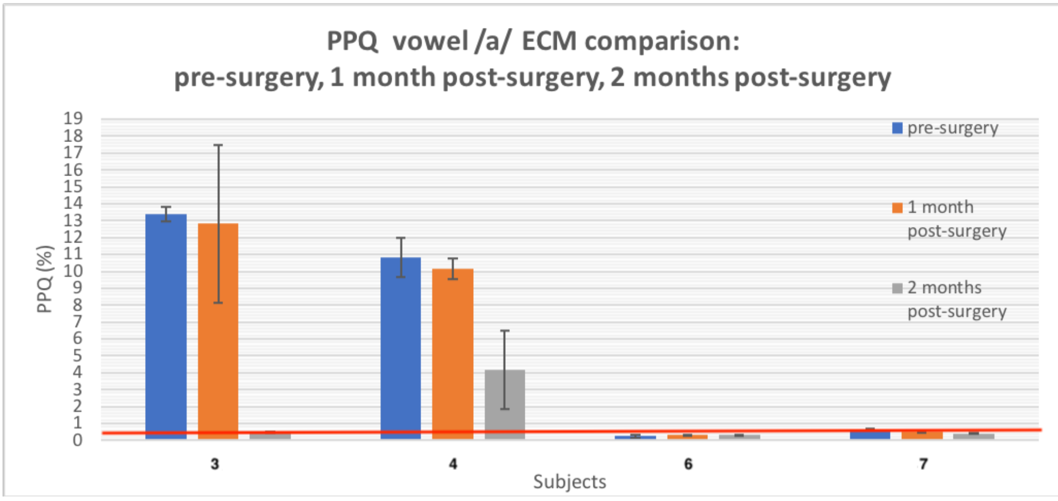

Figure 3.35: Jitt ECM comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean. In red is shown the threshold.

In fig. 3.36 is shown the Shim (%) comparison. All the values obtained are lower than the threshold, so it is not displayed in the histrogram. In all cases there is a lowering of the parameter in exception for the last one.

In fig.3.37 and in fig.3.38 are shown the vAm and HNR parameters. The vAm histogram do not show positive results, in all cases, except for the sixth patient, there is a raise of the parameter value after the operation. The HNR histogram shows a slight increase of the parameter after the operation for all the patients.

In fig.3.39 are reported the CPPS ECM distribution of three vowel /a/ before and after operation of the third and the fourth patient. In all cases, the distribution is moved to higher dB value, evidence of post-operation improvement. The same happens for the sixth patient (fig.3.40), but for the seventh patient the distribution retreats to lower dB values.

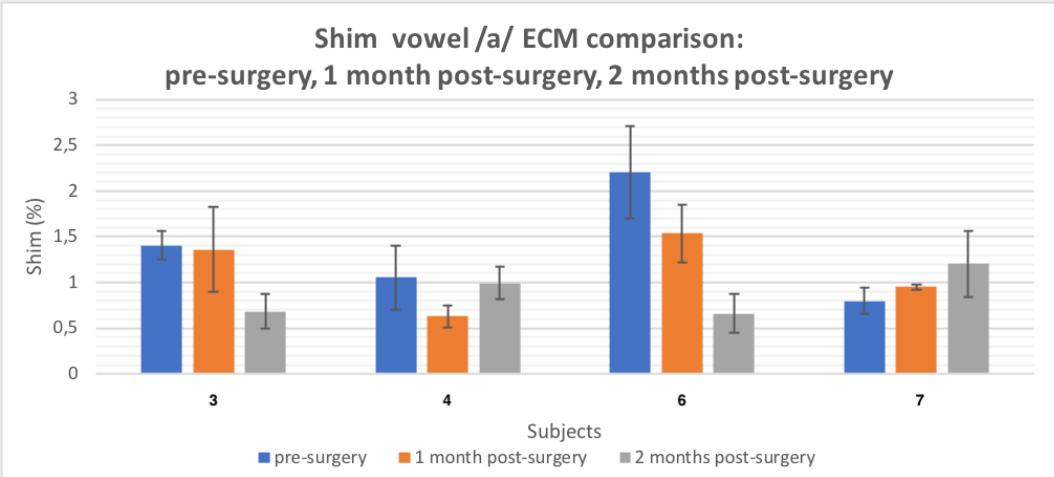

Figure 3.36: Shim ECM comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean.

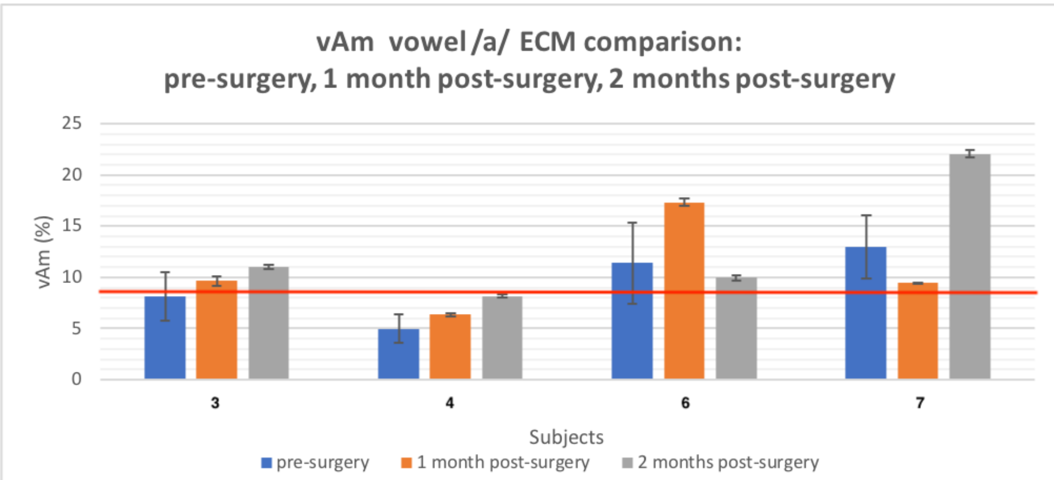

Figure 3.37: vAm ECM comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean. In red is shown the threshold.

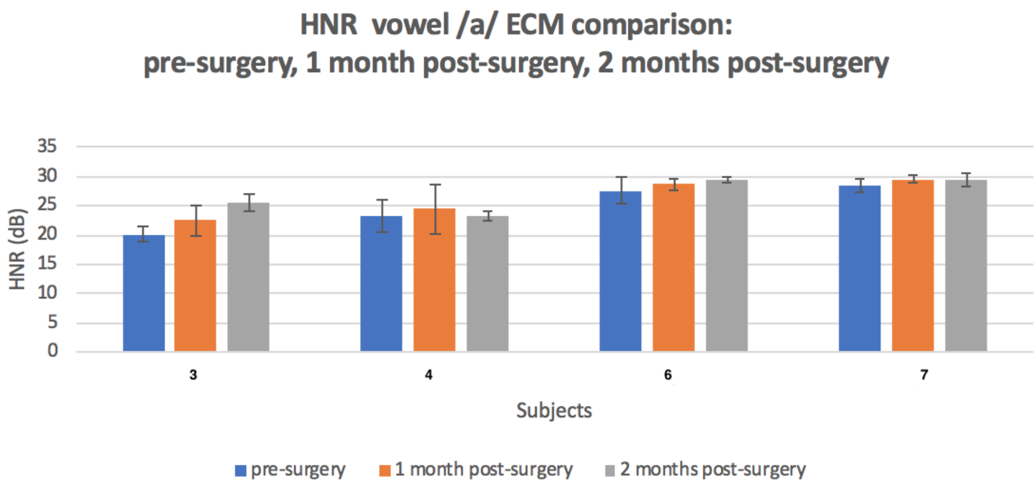

Figure 3.38: HNR ECM comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean.

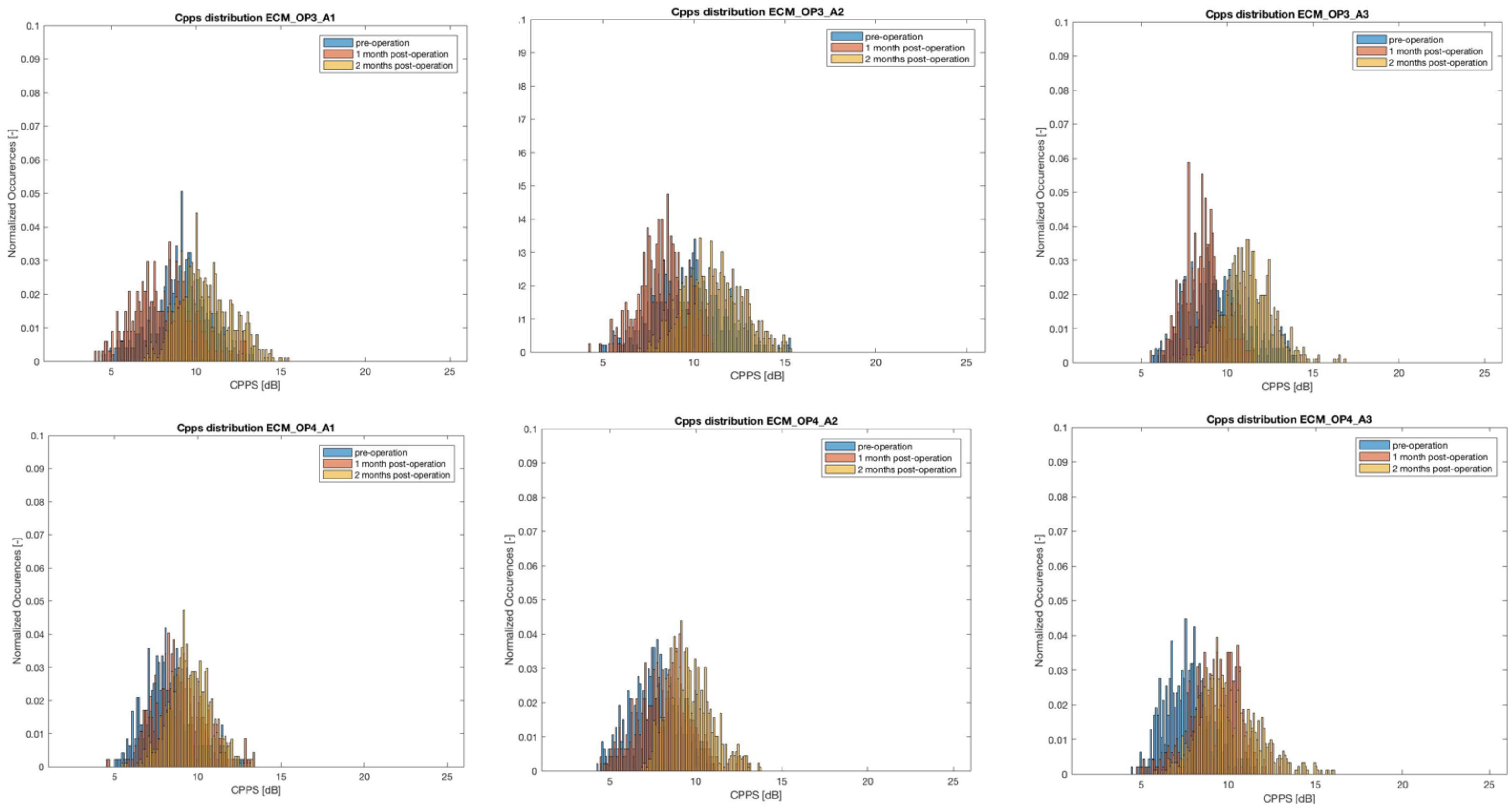

Figure 3.39: CPPS ECM distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery (blue), 1 month post surgery (orange) and 2 months post surgery (yellow). The three graphs above the image belongs to the third patient, the three down belongs to the fourth patient.

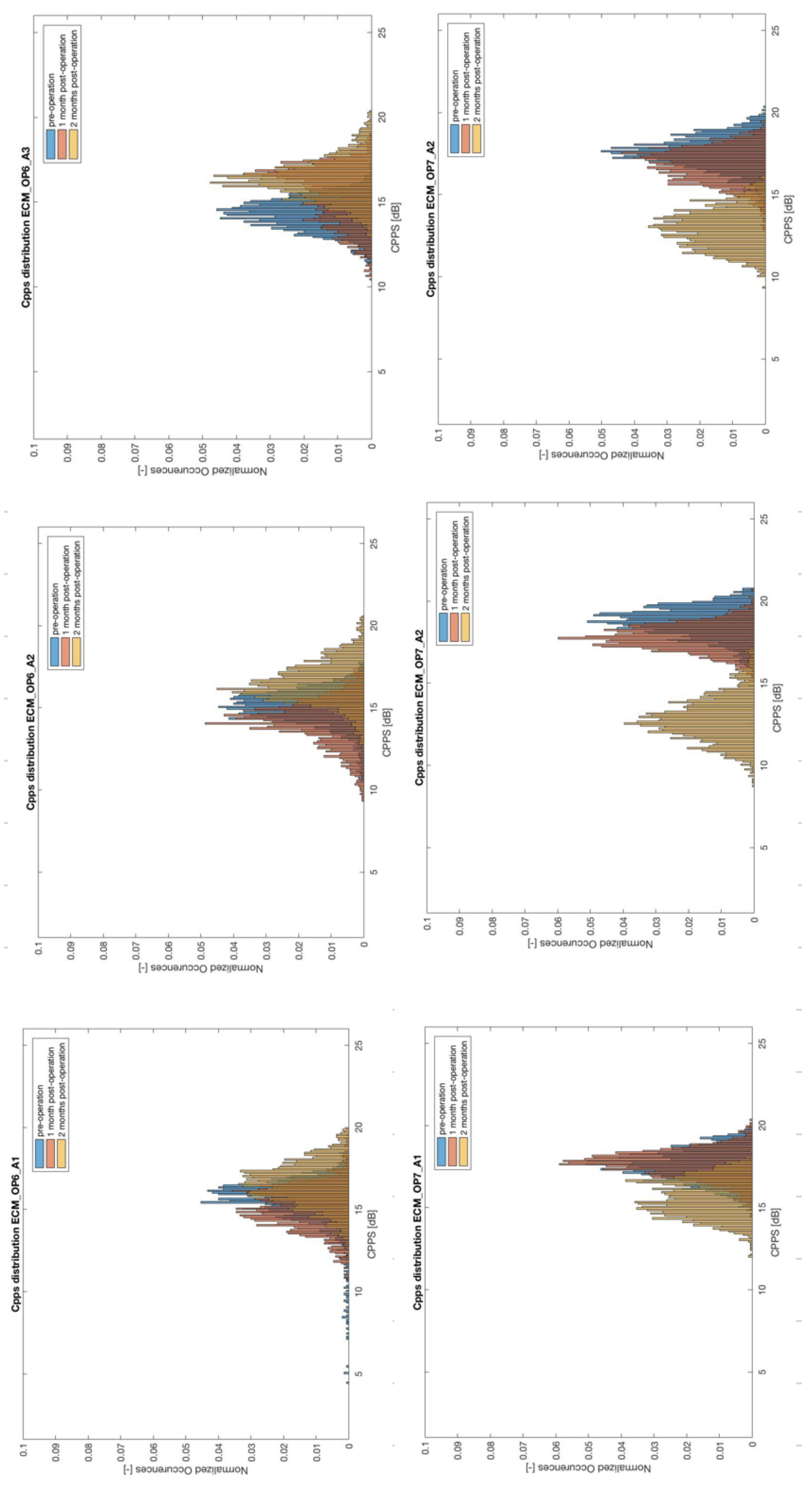

Figure 3.40: CPPS ECM distribution comparison of the three vowel /a/ (A1, A2 and A3) before the surgery (blue), 1 month post surgery (orange) and 2 months post surgery (yellow). The three graphs above the image belongs to the sixth patient, the three down belongs to the seventh patient.

The results seen through the distribution are summarized in fig.3.41, where the CPPS 5prc is shown. In fig.3.42 the CPPSstd histogram shows a lowering below the threshold [13] only for the fourth patient.

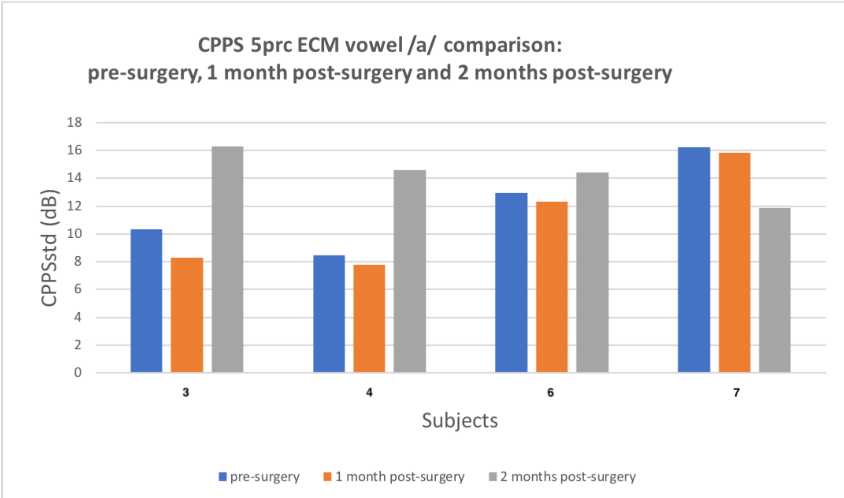

Figure 3.41: CPPS 5prc ECM comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean.

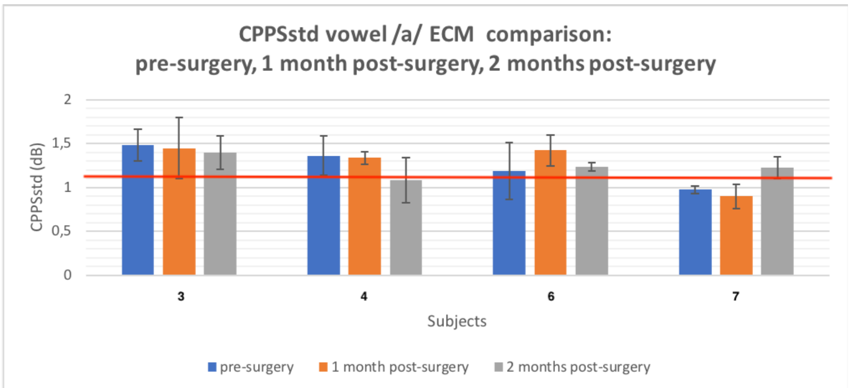

Figure 3.42: CPPSstd ECM comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) vowels /a/ mean.

3.6.6 Mann-Whitney U-test: pre-surgery and post-surgery discrimination

The two-tailed Mann-Whitney U-test is a non-parametric test based on independent samples. It was applied to the subset of parameters related to the two groups: pre-surgery MIPRO and PIEZO vowel /a/ and 2 months post-surgery MIPRO and PIEZO vowel /a/ belonging to all subjects. The null hypothesis (H0) establishes that MD=0, where MD is the median of the population of the differences between the sample data for the two groups. If H0 is accepted (U-test p-value > 0.05), the two lists of values appear to come from the same population. If H0 is rejected (U-test p-value < 0.05), the probability distribution of the relative parameter can be considered enough different for the two groups. The table 3.3 shows that the two-tailed Mann-Whitney U-test p-values for the microphone in air are lower than 0.05 for Jitt, PPQ, Shim and HNR. For the contact microphone PIEZO the H0 is rejected only for Jitt and PPQ. These parameters are the ones who discriminate the best the pre-surgery groups from the post-surgery ones.

In fig.3.43 is shown the Jitt distribution pre-surgery (blue), 1 month post-surgery (orange) and 2 months post surgery (yellow), belonging to MIPRO microphone. It is evident that the yellow bars, belonging to Jitt post-operation values, are shifted more towards low values under the threshold, than those of the pre-operation, the blue ones. The same is evident for the PPQ parameter (fig.3.44). In figure 3.45 and 3.46 are also reported the mean values of the two parameters before and after operation with the corresponding standard errors. In both cases the error bars do not overlap, this means that the two groups are distinct. Instead the vAm parameter has a low discriminatory power that can be seen in the fig.3.47, where the pre and post bars are positioned randomly.

Mann-Whitney U-test p-values		
Parameter	***MIPRO***	***PIEZO***
Jitt	**0.009**	**0.005**
PPQ	**0.000**	**0.000**
Shim	**0.008**	0.125
vAm	0.910	0.153
HNR	**0.021**	0.2818
CPPS5prc	0.185	0.141
CPPSstd	0.430	0.071

Table 3.3: U-test results for each parameter related to MIPRO and PIEZO. Values lower than 0.05 are in bold and indicate the rejection of the null hypotesis.

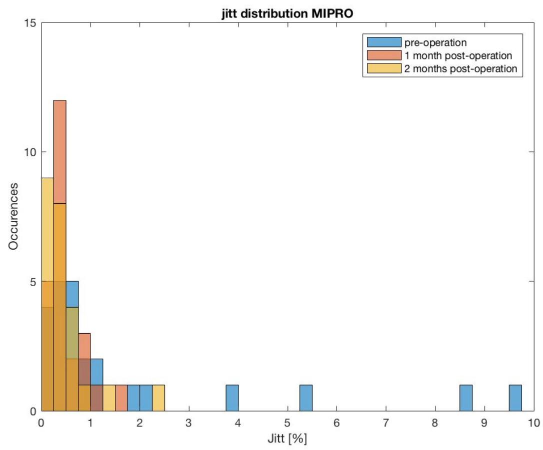

Figure 3.43: Jitt MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow).

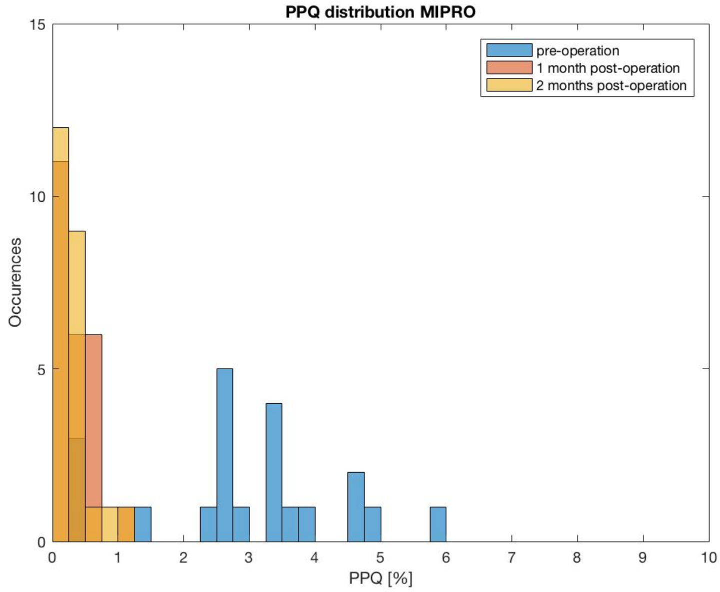

Figure 3.44: PPQ MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow).

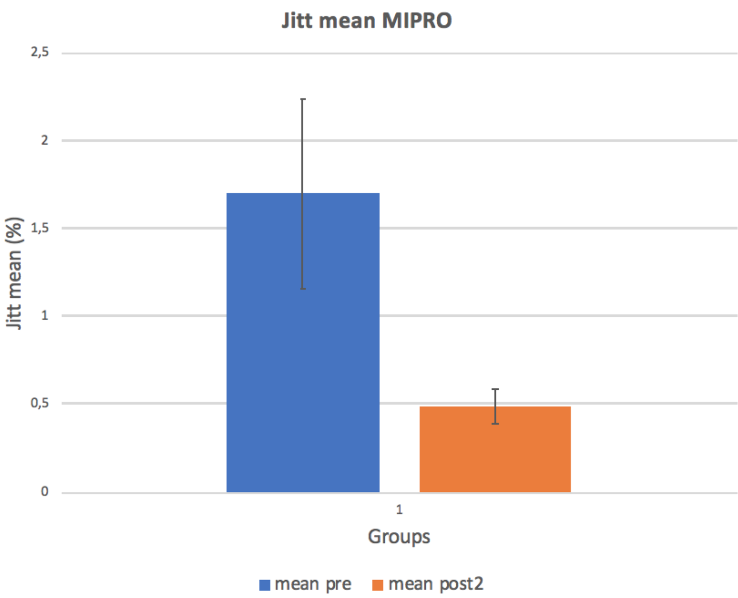

Figure 3.45: Jitt mean MIPRO comparison between pre-surgery (blue) and 2 months post-surgery (orange), for each mean bar the corresponding standard error is plotted.

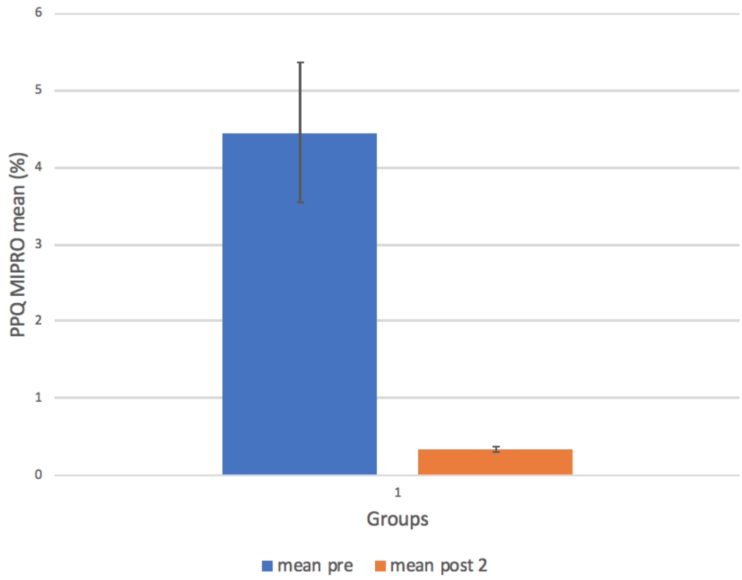

Figure 3.46: PPQ mean MIPRO comparison between pre-surgery (blue) and 2 months post-surgery (orange), for each mean bar the corresponding standard error is plotted.

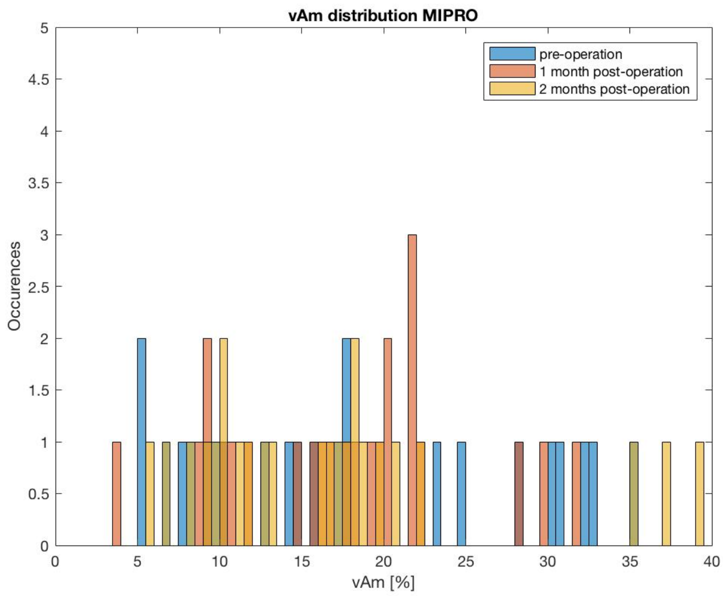

Figure 3.47: vAm MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow).

3.6.7 Microphone comparison

In order to relate the results obtained by the three microphones, the parameters results provided by the MIPRO one was compared to the PIEZO and ECM microphones. In particular the Pearson coefficient (ρ) and Mc Fadden (R^2) were calculated in each comparison (tab.3.4, 3.5). The values obtained for MIPRO-PIEZO comparison are higher than MIPRO-ECM comparison and in each case the Pearson coefficient is higher than R^2 coefficient. The results obtained for Shim parameter by the PIEZO and ECM microphones are very comparable to the MIPRO one. Infact the Pearson coefficient is 0.89 for the comparison MIPRO-PIEZO and 0.71 for the comparison MIPRO-ECM. The parameters HNR, CPPS5th and CPPSstd are similar in the comparison MIPRO-PIEZO, but they do not show an high correlation for the MIPRO-ECM one. Eventually, the Jitt and PPQ comparison do not show good correlation coefficients in both the comparisons.

MIPRO-PIEZO comparison: vowel /a/

Parameters compared	$\boldsymbol{\rho}$	$\boldsymbol{R^2}$
$Jitt_{MIPRO}$-$Jitt_{PIEZO}$	0.57	0.33
PPQ_{MIPRO}-PPQ_{PIEZO}	0.36	0.13
$Shim_{MIPRO}$-$Shim_{PIEZO}$	0.89	0.79
vAm_{MIPRO}-vAm_{PIEZO}	0.41	0.17
HNR_{MIPRO}-HNR_{PIEZO}	0.84	0.71
$CPPS5th_{MIPRO}$-$CPPS5th_{PIEZO}$	0.86	0.74
$CPPSstd_{MIPRO}$-$CPPSstd_{PIEZO}$	0.82	0.68

Table 3.4: MIPRO-PIEZO parametrs comparison. Pearson correlation coefficients (ρ) and Mc Fadden (R^2) are reported.

MIPRO-ECM comparison: vowel /a/		
Parameters compared	$\boldsymbol{\rho}$	$\boldsymbol{R^2}$
$Jitt_{MIPRO}$-$Jitt_{ECM}$	0.32	0.11
PPQ_{MIPRO}-PPQ_{ECM}	0.45	0.20
$Shim_{MIPRO}$-$Shim_{ECM}$	0.71	0.49
vAm_{MIPRO}-vAm_{ECM}	0.17	0.07
HNR_{MIPRO}-HNR_{ECM}	0.59	0.36
$CPPS5th_{MIPRO}$-$CPPS5th_{ECM}$	0.30	0.15
$CPPSstd_{MIPRO}$-$CPPSstd_{ECM}$	0.54	0.29

Table 3.5: MIPRO-PIEZO parametrs comparison. Pearson correlation coefficients (ρ) and Mc Fadden (R^2) are reported.

3.7 Results and Discussion: Reading and Free Speech

The reading and free speech files were analyzed only through the CPPS parameters. In particular, attention was focused on the CPPS 95th percentile and the CPPS std for MIPRO, while on the CPPSmean and on CPPSstd for the PIEZO and the ECM.

3.7.1 Microphone in air: pre-surgery and post-surgery comparison

The same type of graphics described for the vowel /a/ were also plotted for free speech and reading. In fig.3.48, 3.49, fig.3.50, 3.51, fig.3.52, 3.53, fig.3.54, fig.3.55 are shown the CPPS distribution of reading (left graph in the image) and free speech (right graph in the image) from the first to the eighth patient respectively, in the three different monitorings: pre-surgery, 1 month post-surgery and 2 months post-surgery. In all cases there is a high similarity between the CPPS distribution of two tasks (reading and free speech) that is kept in the three different monitorings. In literature there are some evidence about this topic [21], in particular this similarity is called CPPS vocal print.
The distributions belonging from the first to the sixth patient show a movement towards higher dB after the operation, which is representative of an improvement in the health of the vocal apparatus. Only the seventh and the eighth patient show a regression of the distribution after the operation in the reading task and a stationariness in all three monitorings in the free speech task.

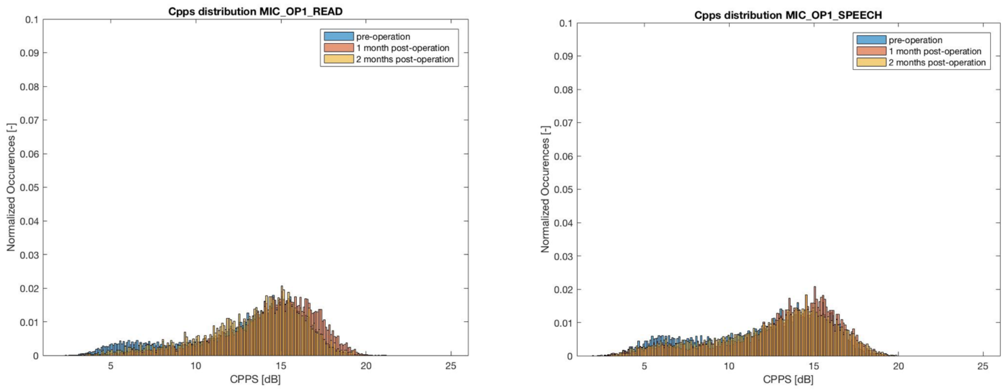

Figure 3.48: CPPS MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the first patient.

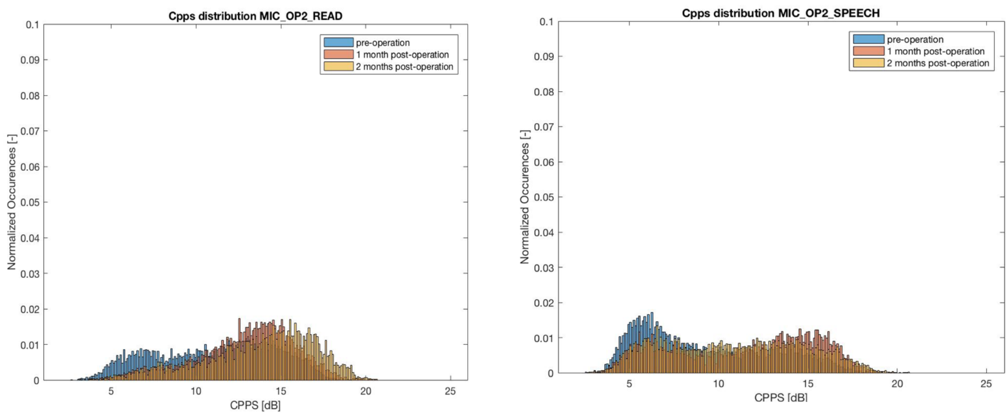

Figure 3.49: CPPS MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the second patient.

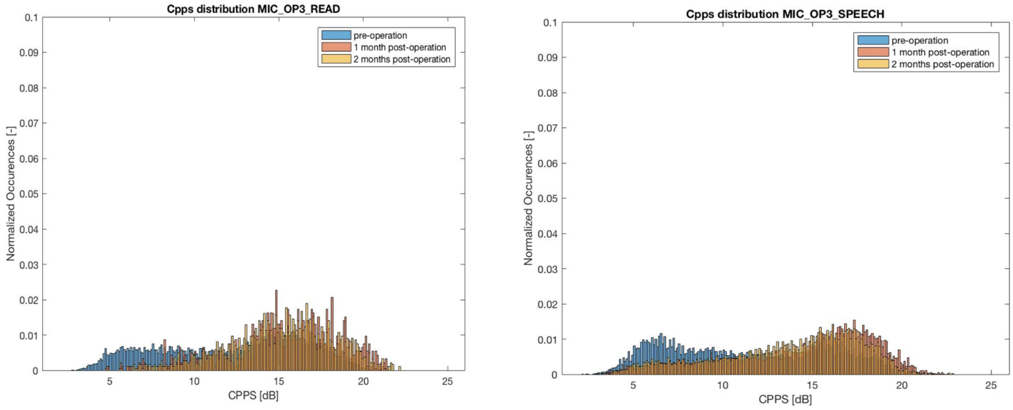

Figure 3.50: CPPS MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the third patient.

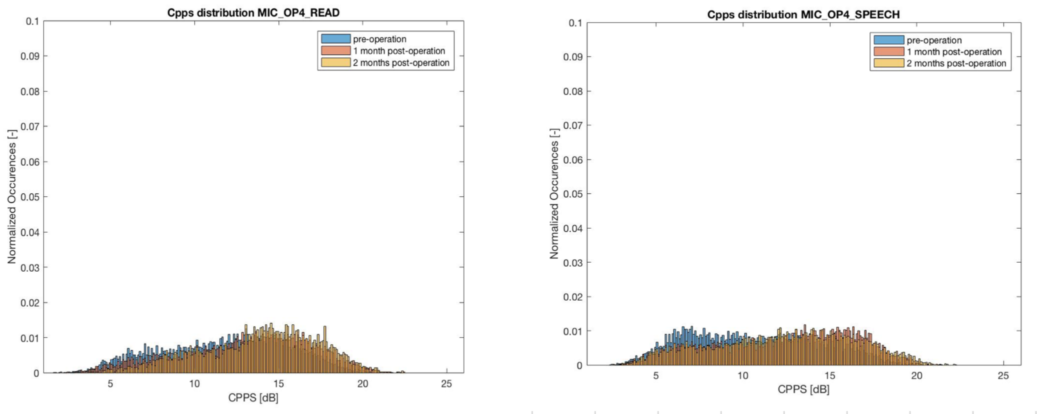

Figure 3.51: CPPS MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the fourth patient.

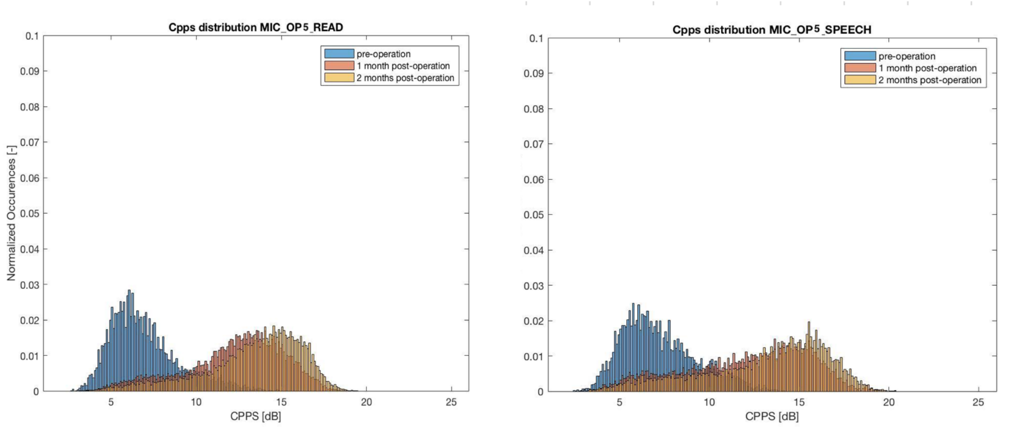

Figure 3.52: CPPS MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the fifth patient.

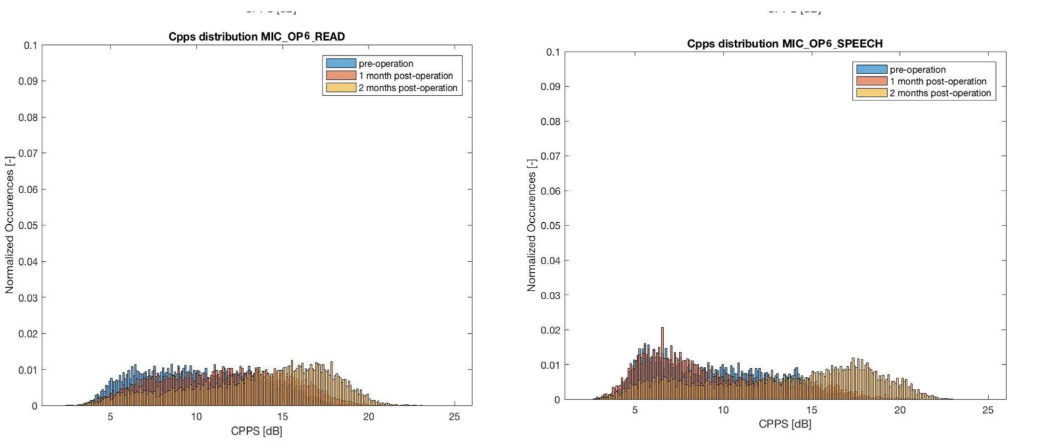

Figure 3.53: CPPS MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the sixth patient.

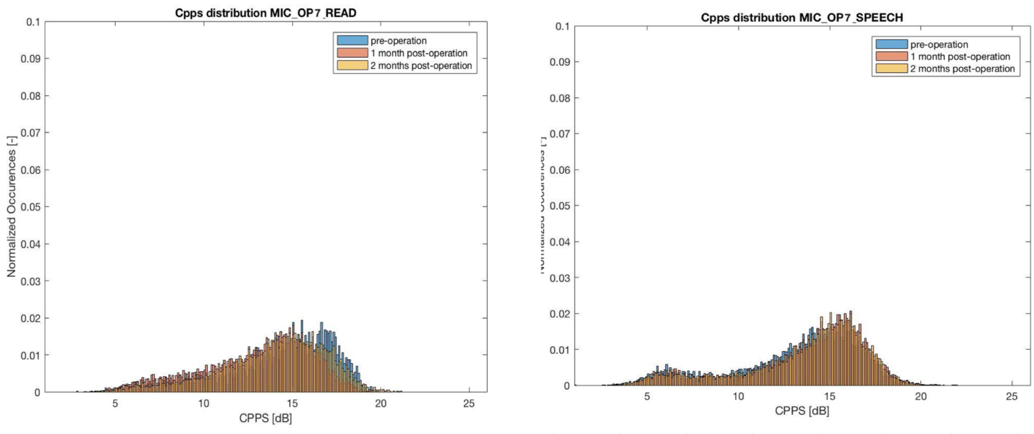

Figure 3.54: CPPS MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the seventh patient.

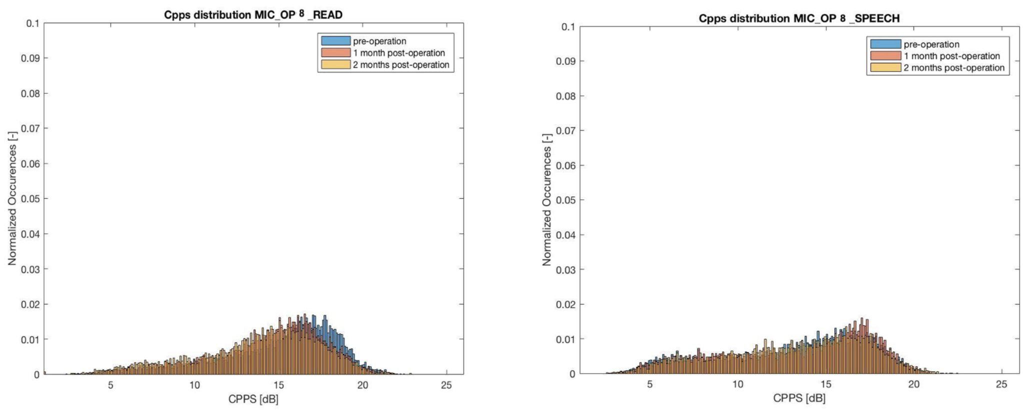

Figure 3.55: CPPS MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the eighth patient.

In fig.3.56 and 3.57 are shown the pre-surgery and post-surgery comparison of CPPS 95prc parameter for reading and free speech respectively. In both the tasks the results are very comparable. In particular for the first to the sixth patients there is a raise of the parameter, only the fifth without going above the threshold in the post-operation. The eigth and the seventh patient parameter do not change in the different monitorings.

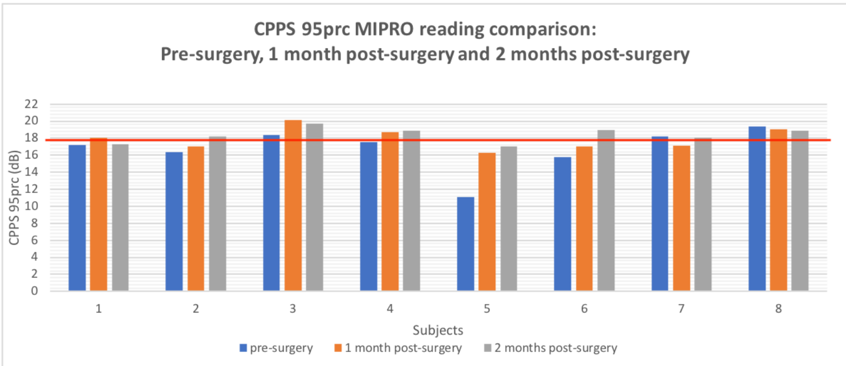

Figure 3.56: CPPS 95prc MIPRO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) reading. In red is shown the threshold for discrimination between healthy and unhealthy voices.

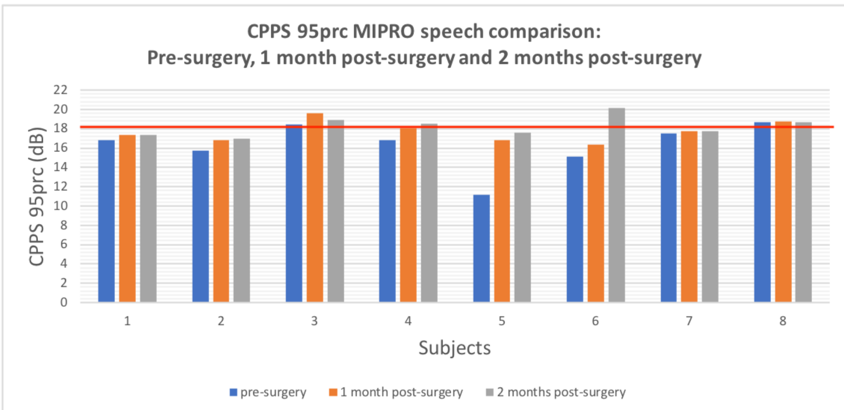

Figure 3.57: CPPS 95prc MIPRO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) free speech. In red is shown the threshold for discrimination between healthy and unhealthy voices.

Eventually, also the CPPSstd parameter shows similarity between the reading task (fig.3.58) and free speech task (fig.3.59). In particular there is a lowering of the parameter for all the patients two months after the operation, except for the fifth and seventh patient.

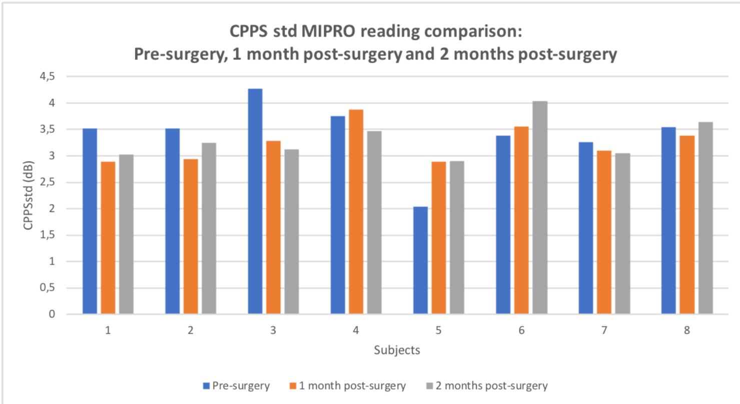

Figure 3.58: CPPSstd MIPRO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) reading. In red is shown the threshold for discrimination between healthy and unhealthy voices.

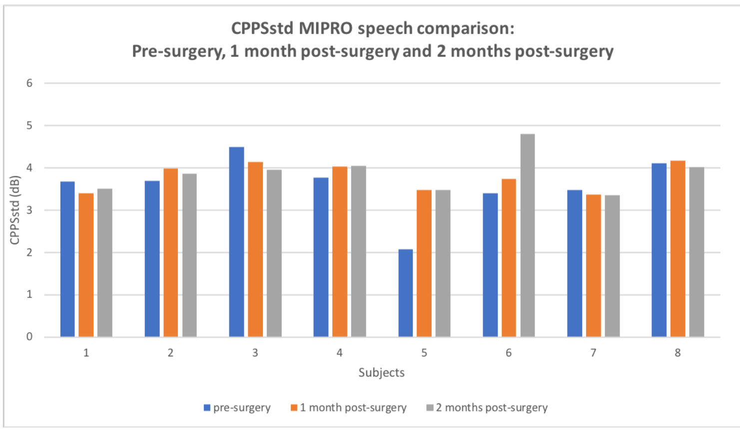

Figure 3.59: CPPSstd MIPRO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (green) free speech. In red is shown the threshold for discrimination between healthy and unhealthy voices.

3.7.2 Piezoelectric contact microphone: pre-surgery and post-surgery comparison

The same evaluations are made considering the results obtained by the PIEZO microphone. Firstly, for each patient the CPPS distribution of the pre-surgery and post-surgery monitorings overlapped are showed (fig.3.60, fig.3.61, fig.3.62, fig.3.63, fig.3.64, fig.3.65, fig.3.66, fig.3.67). Also for the contact microphone, the distribution obtained in the free speech task and reading task are very similar to each other for all the patients. In each case, there is a shift in the distribution to higher dB values, except for the first and the last two patients. This is reflected on the $CPPS_{\mathrm{mean}}$, which is a position parameter, for both reading and free speech tasks (fig.3.68, fig.3.69). Also for the CPPSstd parameters the results obtained are very similar for reading and free speech. The value decreases only for the first, the second and the seventh patient.

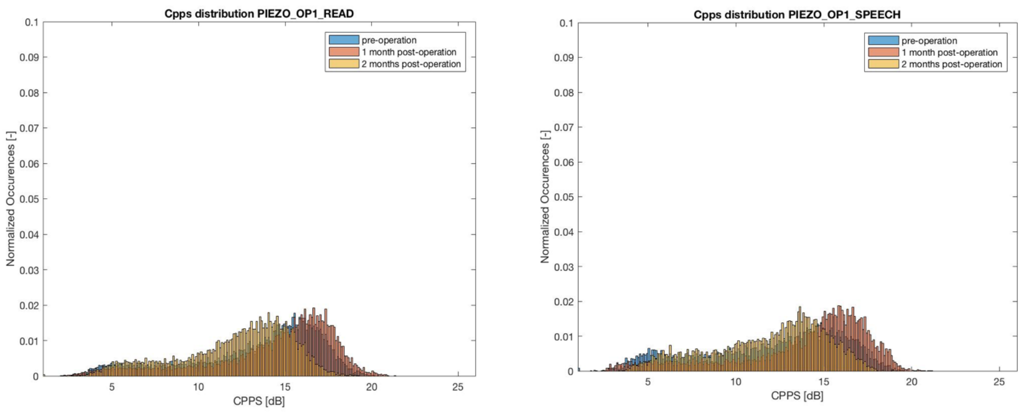

Figure 3.60: CPPS PIEZO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the first patient.

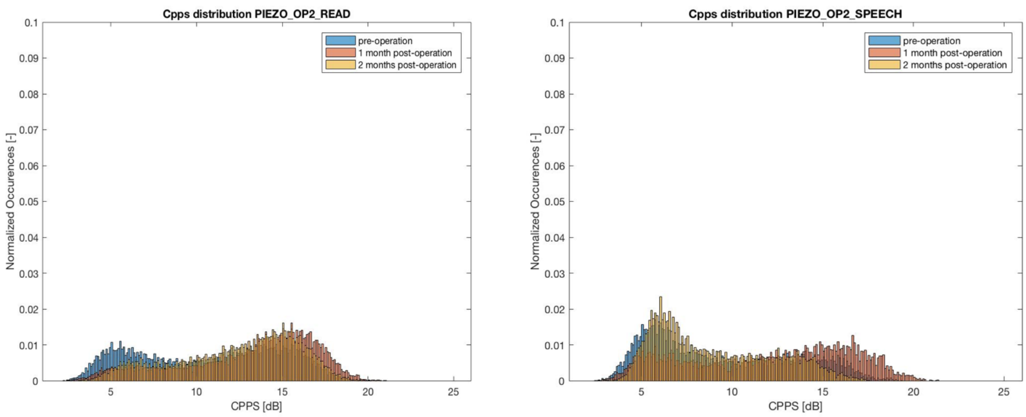

Figure 3.61: CPPS PIEZO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the second patient.

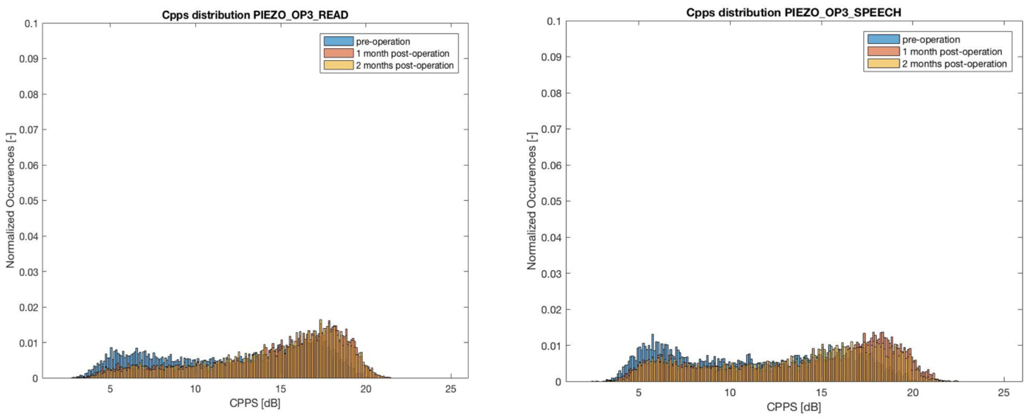

Figure 3.62: CPPS PIEZO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the third patient.

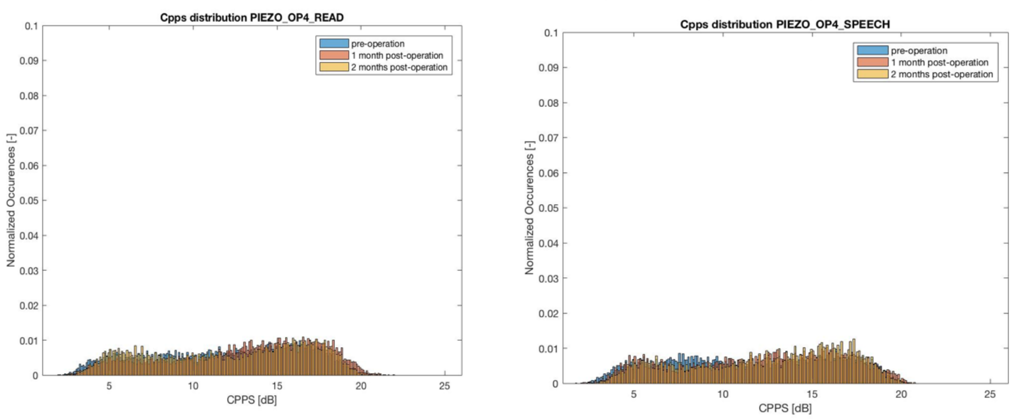

Figure 3.63: CPPS PIEZO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the fourth patient.

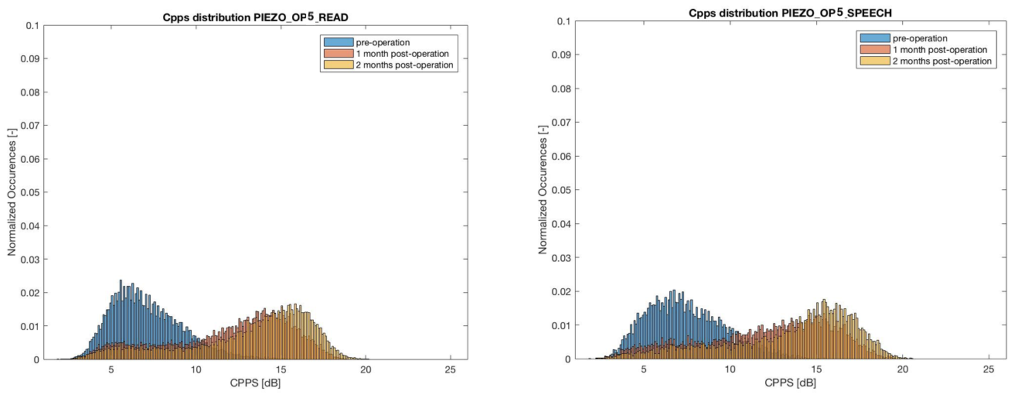

Figure 3.64: CPPS PIEZO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the fifth patient.

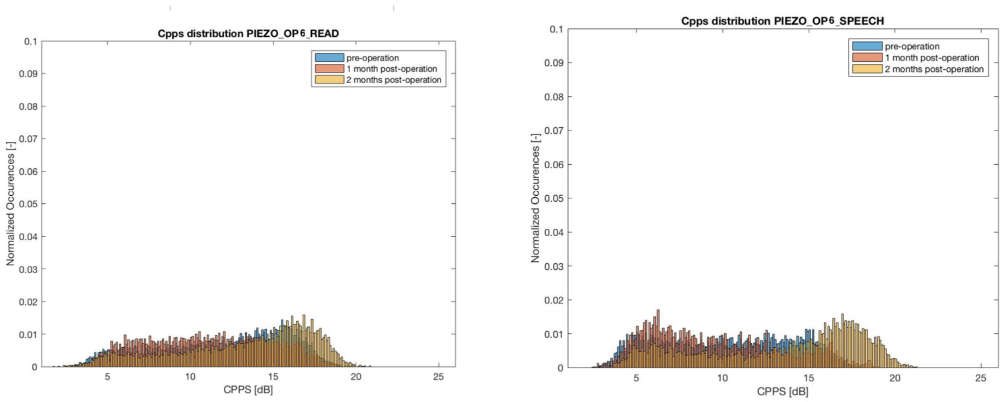

Figure 3.65: CPPS PIEZO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the sixth patient.

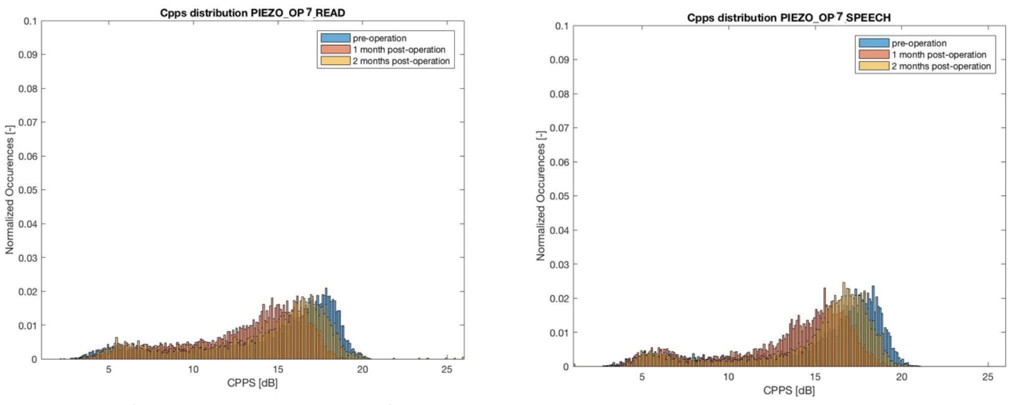

Figure 3.66: CPPS MIPRO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the seventh patient.

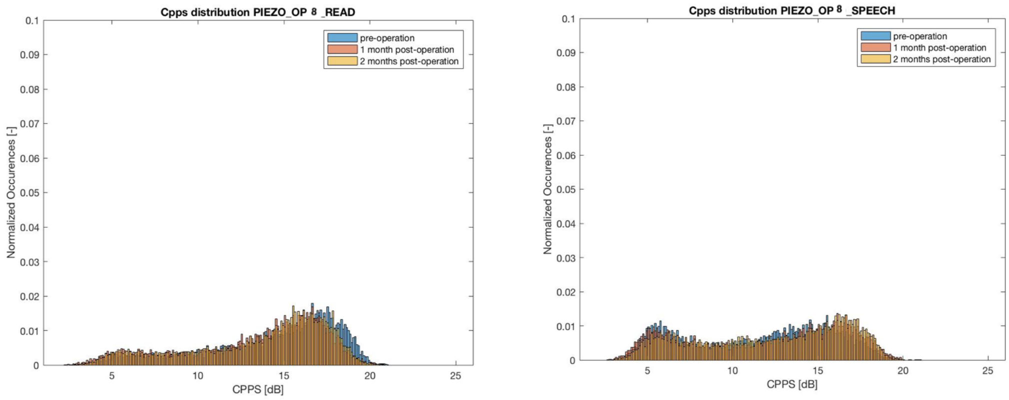

Figure 3.67: CPPS PIEZO distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the eighth patient.

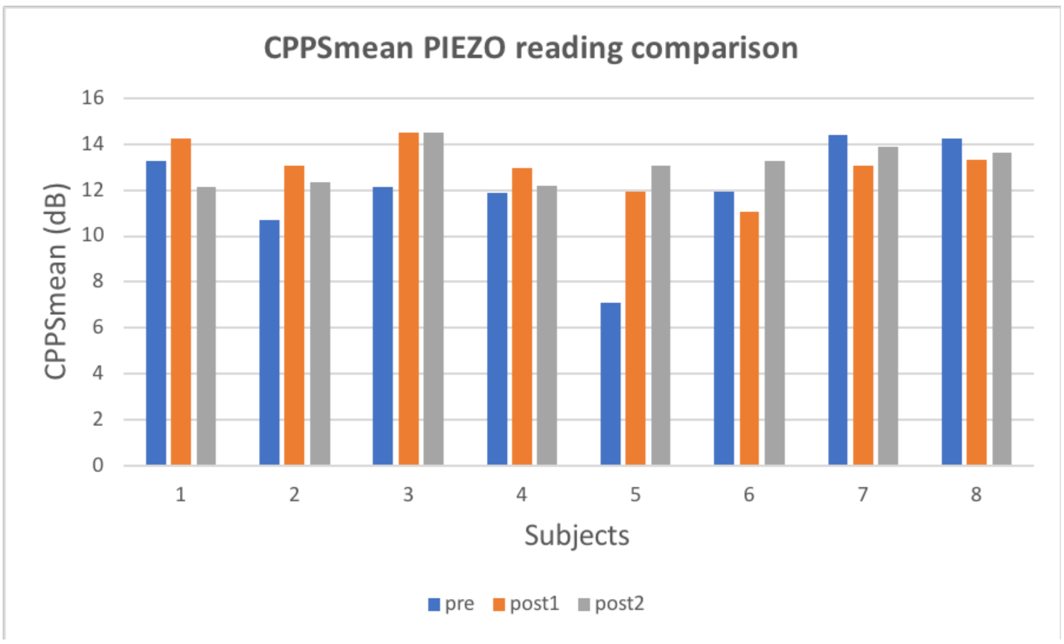

Figure 3.68: CPPSmean PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) reading tasks.

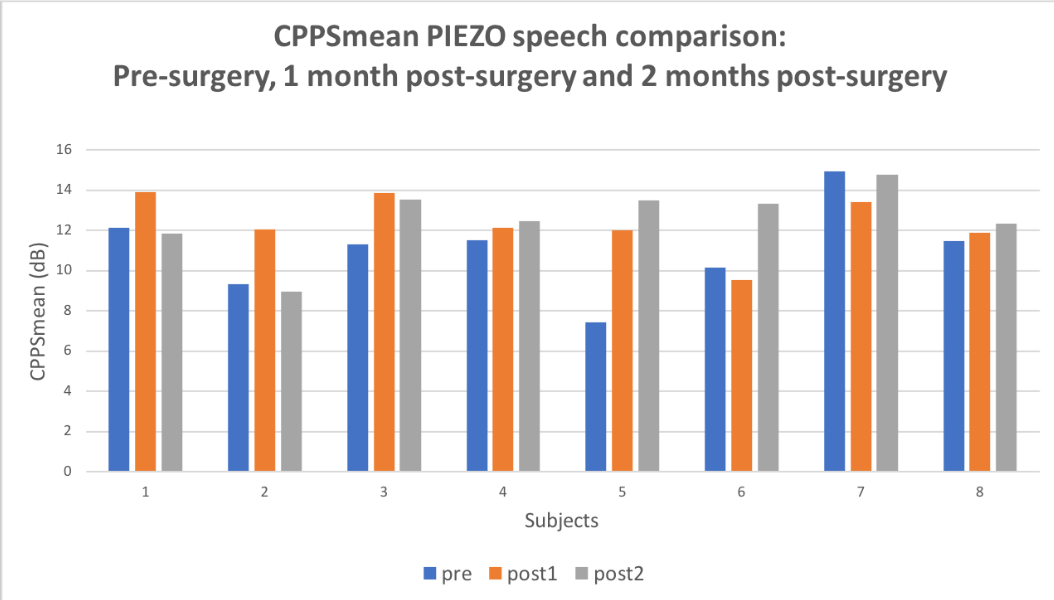

Figure 3.69: CPPSmean PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) free speech tasks.

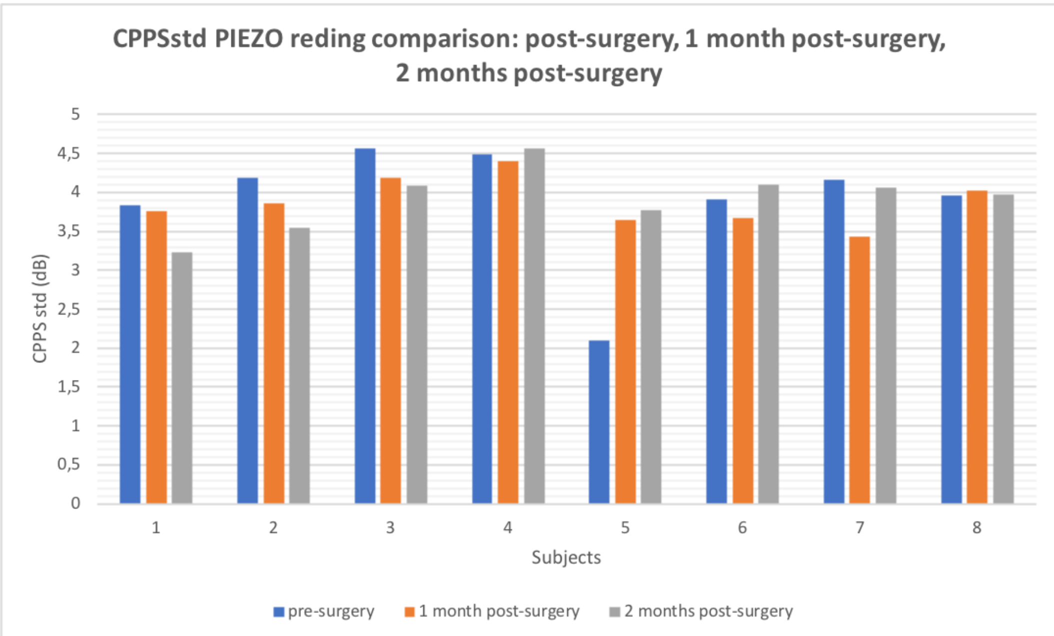

Figure 3.70: CPPSstd PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) reading tasks.

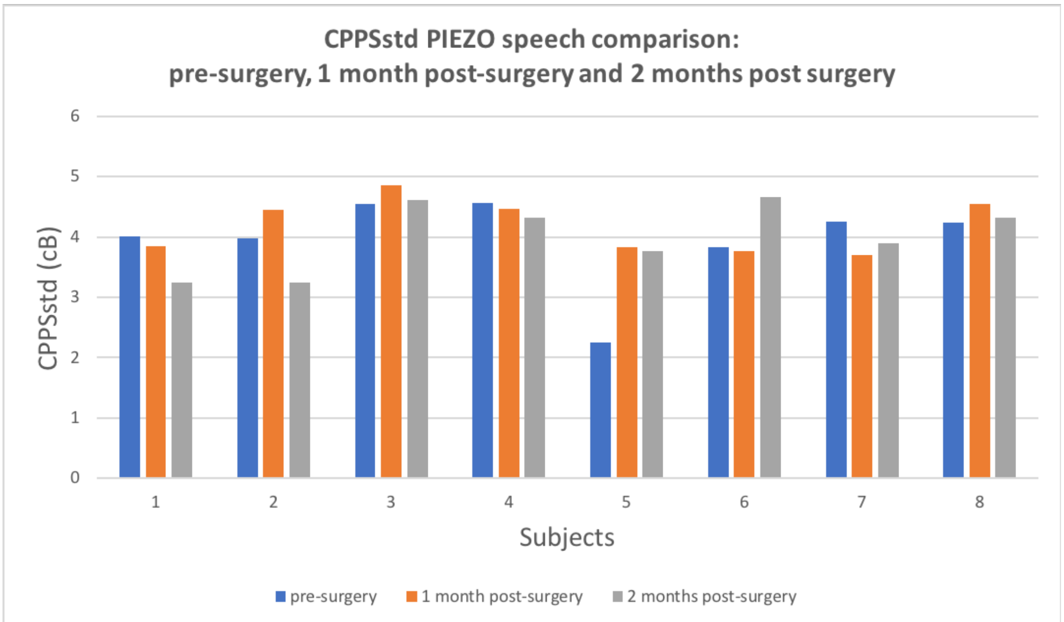

Figure 3.71: CPPstd PIEZO comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) free speech tasks.

3.7.3 Electret Condenser Microphone: pre-surgery and post-surgery comparison

Similar consideration are made for the ECM microphone. In fig.3.72, fig.3.73, fig.3.74 and in fig.3.75 are showed The CPPS ECM distributions comparison of the reading and free speech tasks before and post-operation, for patients three,four,six and seven. For each subject, there is a shape similarity between the reading and free speech CPPS distribution and the one obtained 2 months after the operation is more displaced towards high dB values. This can be notice also through the fig.3.76 and 3.77, that show respectively the CPPSmean parameter comparison for the reading task and the CPPSmean parameter comparison of the free speech task.

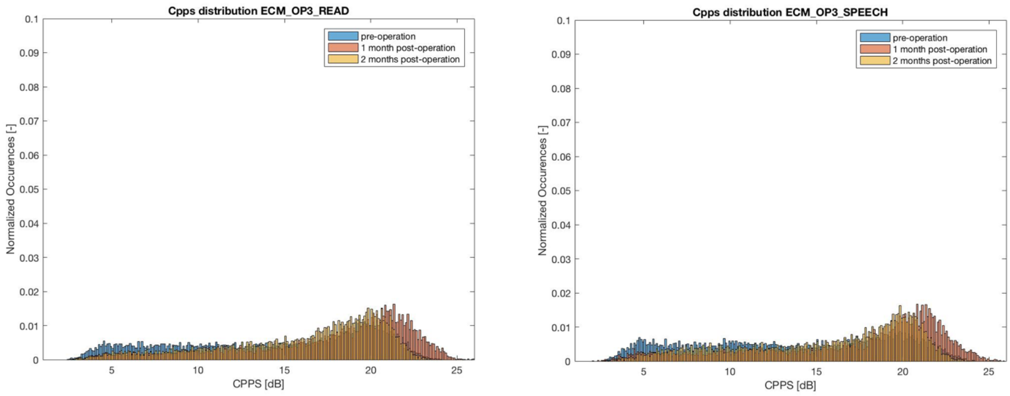

Figure 3.72: CPPS ECM distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the third patient.

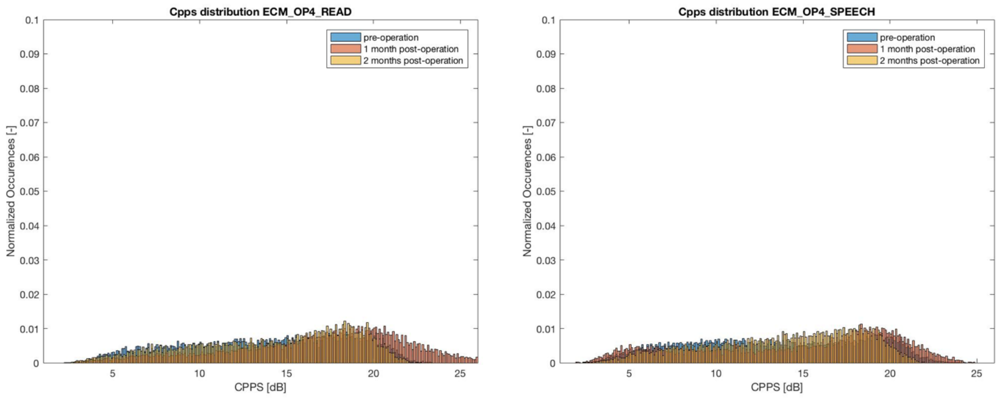

Figure 3.73: CPPS ECM distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the fourth patient.

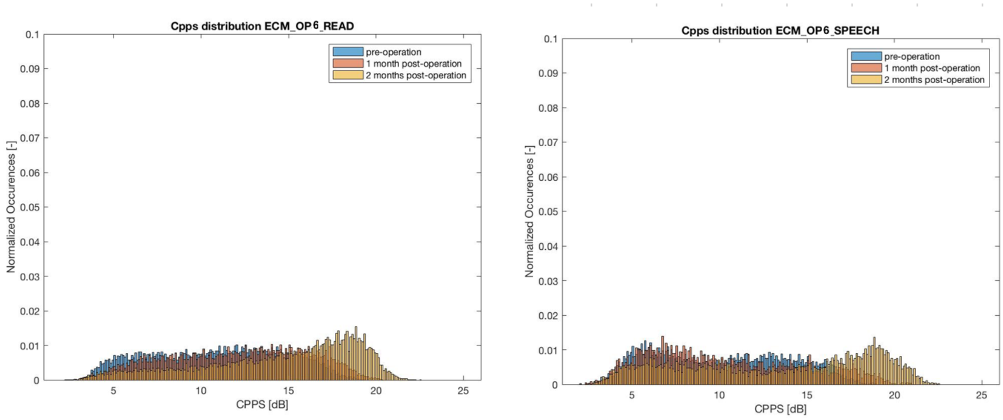

Figure 3.74: CPPS ECM distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the sixth patient.

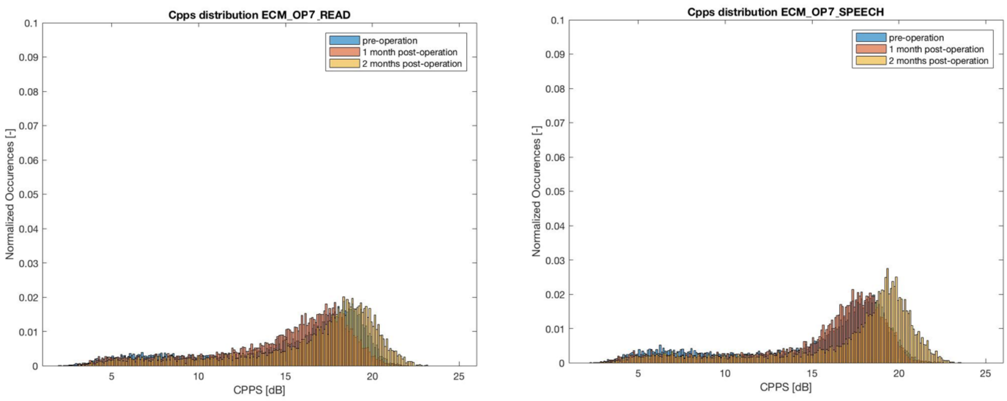

Figure 3.75: CPPS ECM distribution comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading (graph on the left) and free speech (graph on the right) belonging to the seventh patient.

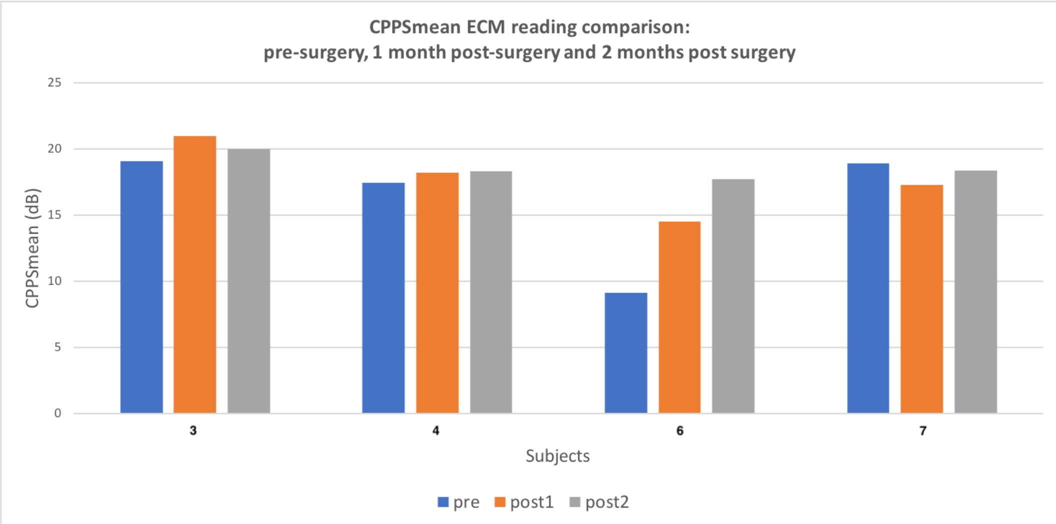

Figure 3.76: CPPSmean ECM comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of reading tasks.

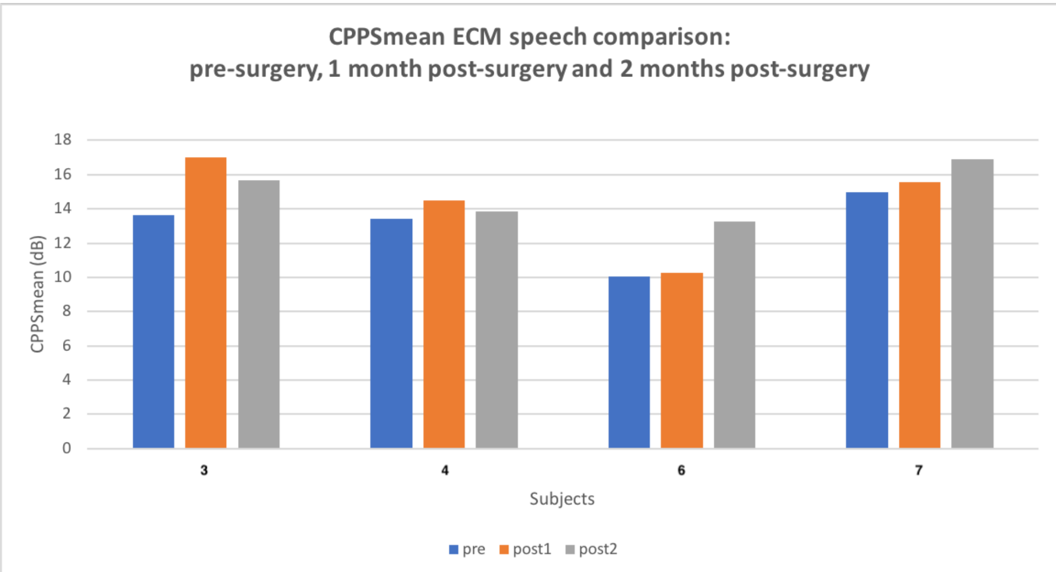

Figure 3.77: CPPSmean ECM comparison between pre-surgery (blue), 1 month post-surgery (orange) and 2 months post-surgery (yellow) of free speech tasks.

3.7.4 Microphone comparison

In order to relate the results of reading and free speech tasks obtained by the three microphones, the parameters results provided by the microphone in air were compared to the PIEZO and ECM ones. In particular the Pearson's coefficient (ρ) and Mc Fadden (R^2) were calculated in each comparison (tab.3.6, 3.7).In general, the values obtained for MIPRO-PIEZO and MIPRO-ECM of reading and free speech are higher than the vowel /a/ values. In particular the MIPRO and PIEZO microphones provide pretty correlated results for almost all the parameters of CPPS in both reading and free speech tasks.

MIPRO-PIEZO comparison

	Reading		Free Speech	
Parameters compared	$\boldsymbol{\rho}$	$\boldsymbol{R^2}$	$\boldsymbol{\rho}$	$\boldsymbol{R^2}$
$CPPSmean_{MIPRO}$-$CPPSmean_{PIEZO}$	0.92	0.86	0.89	0.79
$CPPSmedian_{MIPRO}$-$CPPSmedian_{PIEZO}$	0.91	0.83	0.89	0.80
$CPPSmode_{MIPRO}$-$CPPSmode_{PIEZO}$	0.82	0.68	0.87	0.77
$CPPSrange_{MIPRO}$-$CPPSrange_{PIEZO}$	0.61	0.41	0.60	0.36
$CPPSstd_{MIPRO}$-$CPPSstd_{PIEZO}$	0.78	0.61	0.83	0.69
$CPPS5th_{MIPRO}$-$CPPS5th_{PIEZO}$	0.78	0.61	0.86	0.73
$CPPS95th_{MIPRO}$-$CPPS95th_{PIEZO}$	0.93	0.87	0.86	0.74
$CPPSskew_{MIPRO}$-$CPPSskew_{PIEZO}$	0.92	0.84	0.91	0.84
$CPPSkurt_{MIPRO}$-$CPPSkurt_{PIEZO}$	0.91	0.83	0.89	0.8

Table 3.6: MIPRO-PIEZO parametrs comparison of reading and free speech tasks. Pearson correlation coefficients (ρ) and Mc Fadden (R^2) are reported.

MIPRO-ECM comparison

	Reading		Free Speech	
Parameters compared	$\boldsymbol{\rho}$	$\boldsymbol{R^2}$	$\boldsymbol{\rho}$	$\boldsymbol{R^2}$
$CPPSmean_{MIPRO}$-$CPPSmean_{ECM}$	0.86	0.74	0.92	0.84
$CPPSmedian_{MIPRO}$-$CPPSmedian_{ECM}$	0.91	0.82	0.94	0.88
$CPPSmode_{MIPRO}$-$CPPSmode_{ECM}$	0.71	0.49	0.66	0.43
$CPPSrange_{MIPRO}$-$CPPSrange_{ECM}$	0.29	0.20	0.56	0.31
$CPPSstd_{MIPRO}$-$CPPSstd_{ECM}$	0.58	0.34	0.77	0.59
$CPPS5th_{MIPRO}$-$CPPS5th_{ECM}$	0.68	0.46	0.83	0.69
$CPPS95th_{MIPRO}$-$CPPS95th_{ECM}$	0.81	0.65	0.76	0.58
$CPPSskew_{MIPRO}$-$CPPSskew_{ECM}$	0.92	0.85	0.97	0.95
$CPPSkurt_{MIPRO}$-$CPPSkurt_{ECM}$	0.87	0.76	0.96	0.92

Table 3.7: MIPRO-ECM parameters comparison of reading and free speech tasks. Pearson correlation coefficients (ρ) and Mc Fadden (R^2) are reported.

3.8 PAPV: pre-surgery and post-surgery comparison

The aim of this last analysis is to complete the study considering also the patients self-perception of the vocal problem entity. In fact, the protocol contemplates not only the voice recordings but also the PAPV questionnaire to be filled, which is composed of 28 questions divided in five sections: entity of disease, daily communication, social communication, emotional effects and job effects. In particular, the section of job effects is not considered because only 4 out of 8 subjects are workers, consequently only 24 questions are investigated. For each question 10 points has been attributed for a total score of 240 points. Subjects answer questions by placing a cross (x) on a line. A cross on the extreme left side of the line means that the problem is never present (0 points); a cross on the extreme right of the line means that the problem is serious or always present (10 points); a cross on any point in the line between the two ends indicates that the problem is gradually more serious or less serious (from 9 to 1 points).

Each subject filled in the questionnaire before the operation and two months after the operation. In tab.3.8 are reported the mean points for all patients in each section and than the total score, before and after the operation. There is an evident post-operation decrease in each section and in the total score from 103.4 to 36.4.

Then, the points assigned to each section are normalized by the number of questions that make up the section, in order to make them comparable (fig.3.78). There is a decrease in each section and in particular the bar of emotional effect after the operation is the lowest, which means a considerable emotional improvement.

Mean PAPV points		
	Pre-surgery	**Post-surgery**
Entity of desease	5.71	1.45
Daily communication effects	48.4	23.15
Social communication effects	17.05	4.87
Emotional effects	32.17	6.95
Total (240 points)	103.36	36.42

Table 3.8: PAPV mean points in each section and in total for the eight subjects.

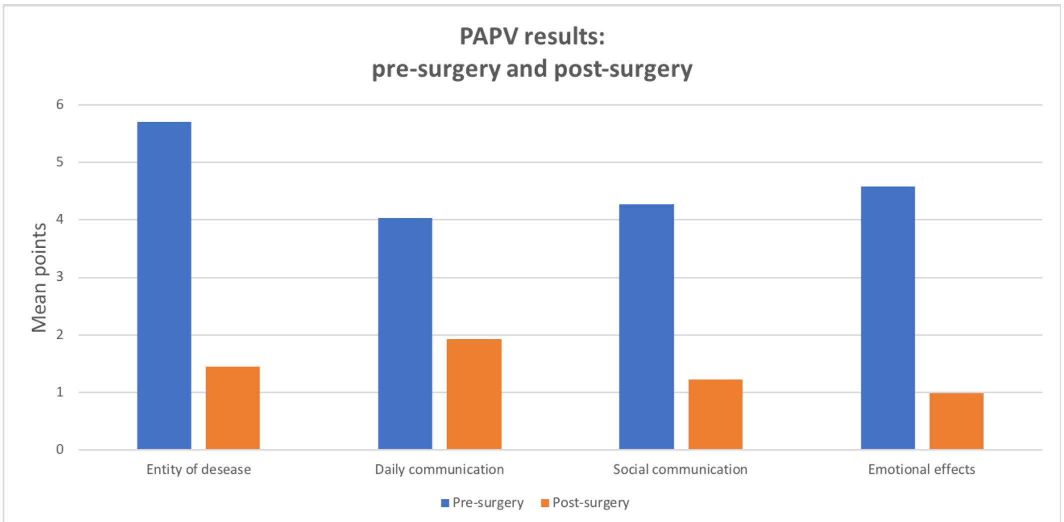

Figure 3.78: Comparison between pre-surgery and post-surgery results of PAPV questionnaire.

Chapter 4

Statistical analysis for discrimination of healthy and pathological voices in sustained vowel /a/

In the previous chapter has been reported the parametric analysis of voice belonging to eight patients who have been subjected to vocal fold surgery. Some parameters have been evaluated and only Jitt, PPQ, Shim and HNR were able to better discriminate the pre-surgery voice than the one after operation. In particular, this result was obtained for the microphone in air vowels /a/.
In this chapter it is described a statistical analysis where these four parameters, in addition to the nine CPPS parameters, are individually evaluated to look for the best discriminator between healthy and pathological voices. Then the best independent parameters are chosen and are coupled to see if the discrimination further improves. This study was carried on for microphone in air vowels /a/.

4.1 Analysis procedure

The subject involved in this study are 40 (20 unhealthy and 20 healthy) which are part of already collected data-base. All the subjects chosen are native Italian speakers and unhealthy ones includes different disorders: vocal fold paresis (5), cyst (3), vocal fold hypostenia (4), vocal fold nodule (3), functional dysponia (3) and chronic laryngitis (2). For each subject the CPPS parameters releated to the first vowel /a/ were collected from the existent database and then the Jitt, PPQ, Shim and vAm parameters were calculated. A table excel were created, where each row belongs to a subject and each column to a parameter. Furthermore, other two columns represent the GIRBAS value G for the unhealthy subjects and a variable that is equal to 1 if the subject is healthy and 0 is the subject is unhealthy, respectively.
A statistical analysis, through the software *RStudio*®, was performed with the 13 parameters. The aim of the statistical analysis was to look for which parameter best discriminates healthy and pathological voices. For this purpose a binary classification analysis, based on the presence of absence of vocal pathology, was performed. The outcome of the videolaringoscopy examination was used as the only reference for checking the vocal health status. A single-variable logistic regression model was performed for each descriptive statistic of CPPS distribution and for Jitt, PPQ, HNR and Shim. The function *gml* of RStudio allows to apply this regression model and it returns the coefficients of the probability functions for each parameter. The algorithm returns the intercept and slope, through which is possible to define the empirical fitted model:

$$P(Unhealthy) = \frac{e^{intercept - slope \cdot parameter}}{1 + e^{intercept - slope \cdot parameter}}$$

The best model was chosen through the evaluation of the Akaike Information Criterion (AIC). The AIC estimates the quality of a parameter in relation with the other parameters. It takes into account the lost information when a model is used, respect to the complexity of the model. Given the AIC values of all the parameters, the best model is the one with lower value of AIC. In addition, the Mc Fadden's R^2 value is calculated. It gives an estimation of the goodness of the logistic regression model. The more the parameter is close to 1, the better the model is. In order to evaluate the differentiation of the two groups (healthy and unhealthy) the Mann-Whitney U-test is executed. The null hypothesis of the test establishes that the groups belongs to the same population and their probability distribution is the same. If the p-value of the test is less than 0.05 the null hypothesis is rejected and the probability distribution of that parameter can be considered enough different for the two groups. So the parameters with lower p-value are the best in discrimination between the two categories.

Moreover the ROC curve is estimated from the probability distributions of healthy and unhealthy groups. The axis of the curve are the sensitivity (x-axis) and 1- specificity (y-axis). The sensitivity is the probability to identify the pathological voice correctly and the specificity is the probability to identify the healthy voices correctly. The Area Under Curve (AUC) is the area under the ROC curve. If AUC is closer or equal to 0.5 it identifies a low capability to discriminate healthy and unhealthy because the probability to classify a subject 0 or 1 is the same. The more the value of AUC is close to 1 the more the classification is correct, because the false positive rate is low and the true positive rate is high. AUC values lower than 0.8 are not considered acceptable for diagnostic purpose. To give more information, also the accuracy is calculated and the leave one out method is implemented. Major accuracy means better prediction.

The optimal threshold for the classification purpose is chosen in order to have the same values of sensitivity (true positive rate) and specificity (true negative rate) giving priority to higher sensitivity.

After that the single-variable logistic regression model was performed, the Pearson's correlation parameter was calculated between all the parameters and the best parameters not correlated where chosen. A double-variable logistic regression model is performed to see if the parameters together let the discriminatory power further increase. Then the analysis is repeated with three parameters together.

4.2 Results and Discussion

The table 4.1 shows the performance on classifying healthy and unhealthy voice of each model with a parameter at a time as independent variable and the presence/absence of voice disorders as dependent variable. The logistic regression model that showed the best capability in discriminating healthy from pathological voice was the one which uses the parameter PPQ as the independent variable with the highest AUC and accuracy. The respective values are 0,89 and 85%. Also the logistic regression model that uses the CPPS 5prc showed good performance with highest R^2 equals to 0,41 and lowest AIC equals to 36,8. The best results are highlighted in bold. The ROC curve of the best logistic models are showed in fig.4.1 and fig.4.2. The empirical fitted model which uses the CPPS5prc has the following expression:

$$P(Unhealthy) = \frac{e^{10.655-0.89 \cdot CPPS5prc}}{1+e^{10.655-0.89 \cdot CPPS5prc}}$$

Where the intercepts is equal to 10.655 and the slope is equal to 0.89. The empirical fitted model which uses the PPQ has the following expression:

$$P(Unhealthy) = \frac{e^{intercept-slope \cdot PPQ}}{1+e^{intercept-slope \cdot PPQ}}$$

Where the intercept value is -1.5961 and the slope value is 1.1980.

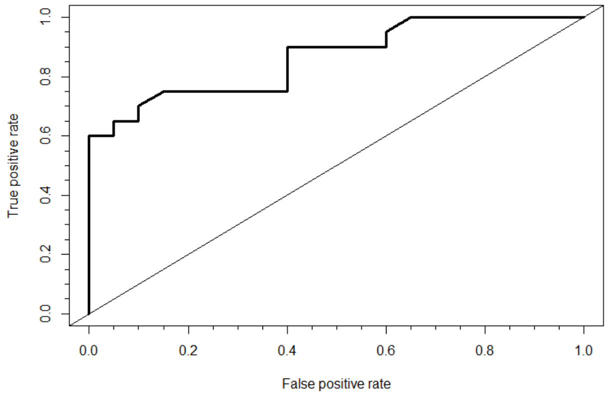

Figure 4.1: ROC curve of the CPPS 5prc logistic model.

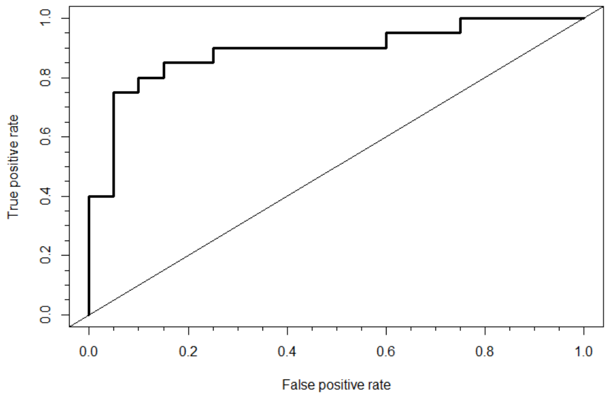

Figure 4.2: ROC curve of the PPQ logistic model.

Parameter	**AIC**	R^2	**U-test**	**AUC**	**ACC**
CPPSmean	37.6	0.39	0.000	0.86	84%
CPPSmedian	38.8	0.37	0.000	0.85	83%
CPPSmode	43.56	0.29	0.001	0.79	81%
CPPSrange	55.69	0.07	0.007	0.66	78%
CPPSstd	49.8	0.17	0.015	0.72	78%
CPPS5prc	**36.8**	**0.41**	**0.000**	**0.87**	**84%**
CPPS95prc	39.42	0.36	0.000	0.86	83%
CPPSskew	59.37	0.0014	0.438	0.57	72%
CPPSkurt	58.93	0.009	0.515	0.56	73%
Jitt	48.65	0.36	0.001	0.79	78%
PPQ	**39.16**	**0.36**	**0.000**	**0.89**	**85%**
Shim	53.08	0.11	0.000	0.85	78%
HNR	56.30	0.056	0.12	0.645	75%

Table 4.1: Analysis results for each parameter related to the microphone in air vowel /a/. The columns represents respectively: AIC, Mc Fadden R^2, Mann-Whitney U-test p-values, Area Under Curve (AUC), leave one out classification accuracy (ACC). The parameters which are used as independent variables of the best logistic regression models are in bold.

Then, the Pearson correlation coefficient was calculated in order to see the correlation between the parameters. In tab.4.3 are shown the results. The boxes highlighted show that the CPPS5thprc and Jitt, PPQ, Shim and HNR are not related. Consequently, a double-variable logistic regression model was performed for the couples: CPPS 5prc-Jitt, CPPS 5prc-PPQ, CPPS 5prc-Shim and CPPS 5prc-HNR. The results are shown in the tab.4.2.

Pearson's correlation coefficient (ρ) **between vowel /a/ parameters**

vowMIPRO	CPPS mean	CPPS median	CPPS mode	CPPS range	CPPS Std	CPPS 5prc	CPPS 95prc	CPPS Skew	CPPS Kurt	Jitt	PPQ	Shim	HNR
CPPS mean		0,99	0,97	-0,46	-0,29	0,96	0,94	-0,49	-0,06	-0,68	-0,54	-0,50	0,59
CPPS Median			0,98	-0,48	-0,3	0,95	0,92	-0,53	-0,03	-0,70	-0,54	-0,51	-0,58
CPPS mode				-0,54	-0,32	0,94	0,87	-0,56	-0,02	-0,74	-0,54	-0,51	0,57
CPPS range					0,81	-0,65	-0,16	-0,01	0,03	0,69	0,52	0,46	-0,25
CPPS std						-0,49	-0,05	-0,28	0,49	0,42	0,26	0,33	-0,10
CPPS 5prc							0,84	-0,30	-0,12	-0,62	-0,57	-0,54	0,56
CPPS 95prc								-0,46	-0,08	-0,50	-0,42	-0,4	0,57
CPPS Skew									-0,38	0,37	0,26	0,24	-0,3
CPPS Kurt										-0,12	-0,16	-0,02	0,01
Jitt											0,71	0,78	-0,57
PPQ												0,57	-0,43
Shim													-0,42
HNR													

Figure 4.3: Pearson correlation coefficient between all the parameters. The boxes highlighted show that the CPPS 5th prc and Jitt, PPQ, Shim and HNR are not related

Parameter	AIC	R^2	AUC	ACC
CPPS5prc+Jitt	38.63	0.41	0.86	83%
CPPS5prc+PPQ	**31.41**	**0.54**	**0.93**	**86%**
CPPS5prc+Shim	38.25	0.42	0.87	82%
CPPS5prc+HNR	38.48	0.41	0.87	83%

Table 4.2: Analysis results for each parameter related to the microphone in air vowel /a/. The columns represents respectively: AIC, Mc Fadden R^2, Area Under Curve (AUC), leave one out classification accuracy (ACC). The couple of parameters which are used as independent variable of the best logistic regression model are in bold.

The parameters that individually showed the best performance as independent variables in the single-variable logistic models were coupled and showed a better capability in in discriminating healthy and pathological voices as dependent variables of the double-variable logistic model. The results showed AUC equal to 0.93 and an accuracy of 86% and both are higher than the one obtained for PPQ and CPPS5prc individually. The ROC curve is represented in figure 4.4.

The empirical fitted model in this case has the following expression:

$$P(Unhealthy) = \frac{e^{12.202-1.105\cdot CPPS5prc+1.278\cdot PPQ}}{1+e^{12.202-1.105\cdot CPPS5prc+1.278\cdot PPQ}}$$

Where the intercept is equal to 12.202 the slope for CPPS5prc parameter is -1.105 and the slope for PPQ parameter is 1.278.

Eventually, the logistic regression was carried out with three parameters together: CPPS5prc , PPQ and Shim. Complicating the model no better results are obtained, in fact for the AUC and the accuracy are obtained 0,93 and 86%, respectively.

The cut off values, relative to the best logistic models,were evaluated in order to see the thresholds that best discriminates healthy and unhealthy subjects

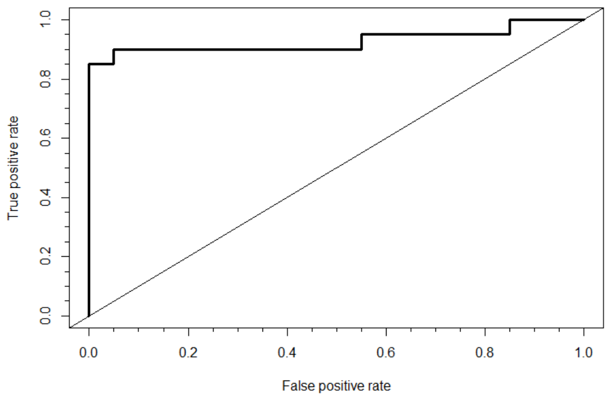

Figure 4.4: ROC curve of the CPPS5prc+PPQ logistic model.

for CPPS 5prc, PPQ and for CPPS5prc+PPQ models (tab.4.3). In figure 4.5,

Parameter	**SENS**	**SPEC**	**Cut off**
CPPS5prc	0.80	0.85	0.51
PPQ	0.85	0.85	0.36
CPPS5prc+PPQ	0.90	0.95	0.41

Table 4.3: Threshold (cut off) and respective sensitivity (SENS) and specificity (SPEC) of the best logistic models.

4.6 and 4.7 are showed the specificity and sensitivity in function of the cutoff for the three models, the thresholds are chosen in correspondence of the cross between the two probabilities.

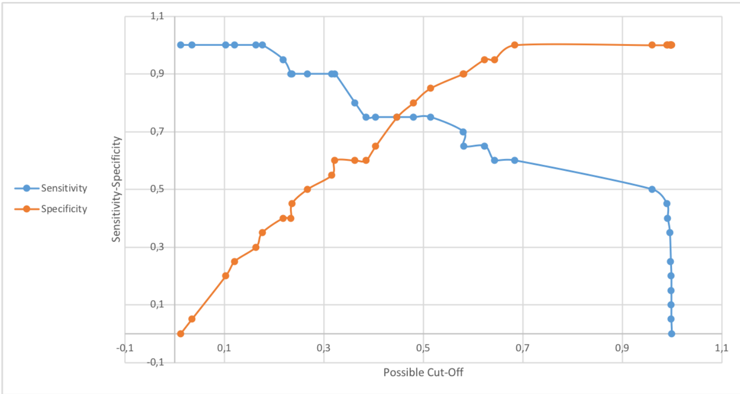

Figure 4.5: Sensitivity and Specificity at the variation of the cut-off for CPPS5prc.

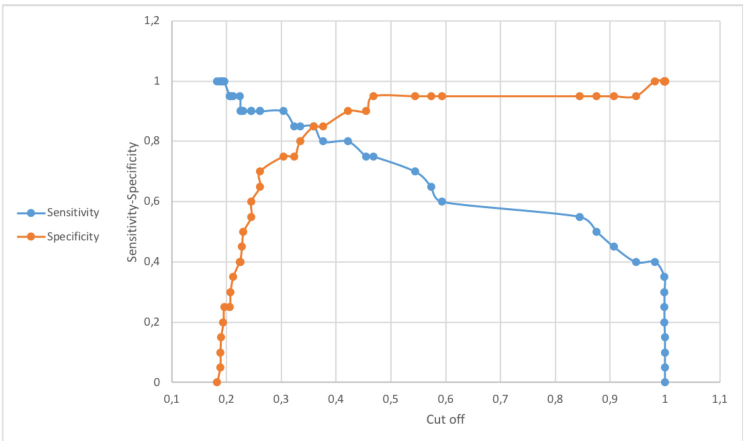

Figure 4.6: Sensitivity and Specificity at the variation of the cut-off for PPQ.

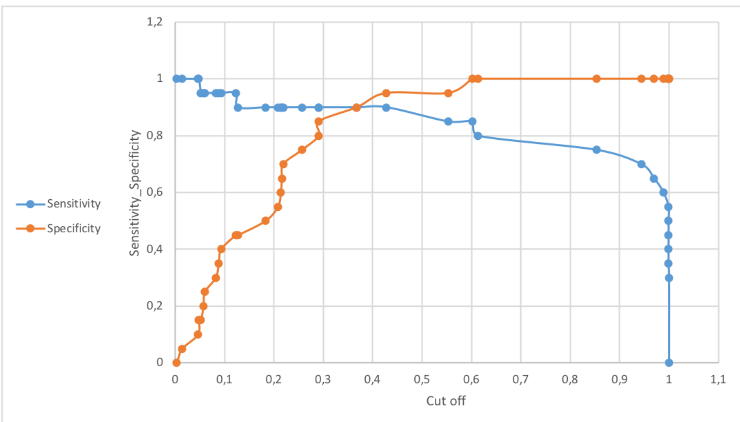

Figure 4.7: Sensitivity and Specificity at the variation of the cut-off for CPPS5prc+PPQ.

The figure 4.8 shows the values provided by the logistic model, with the CPPS 5prc like independent variable, for each subject. Most of the pathological patients (circle points) are in the upper part of the graph where the probability of having unhealthy voice is near to one, while most of the healthy patients (cross points) have lower scores close to 0, with only 2 patients classified incorrectly. Furthermore there is a partial agreement between the overall grade of dysphonia G and the probability of having unhealthy voice: all patients with G=3 are close to score 1, patients with G=2 are above the threshold with only one miss classified, and subjects with G=1 have four incorrectly classified because they were indicated with the lowest dysphonia rate. The figure also shows the best classification threshold in terms of P(Unhealthy)=0.51.

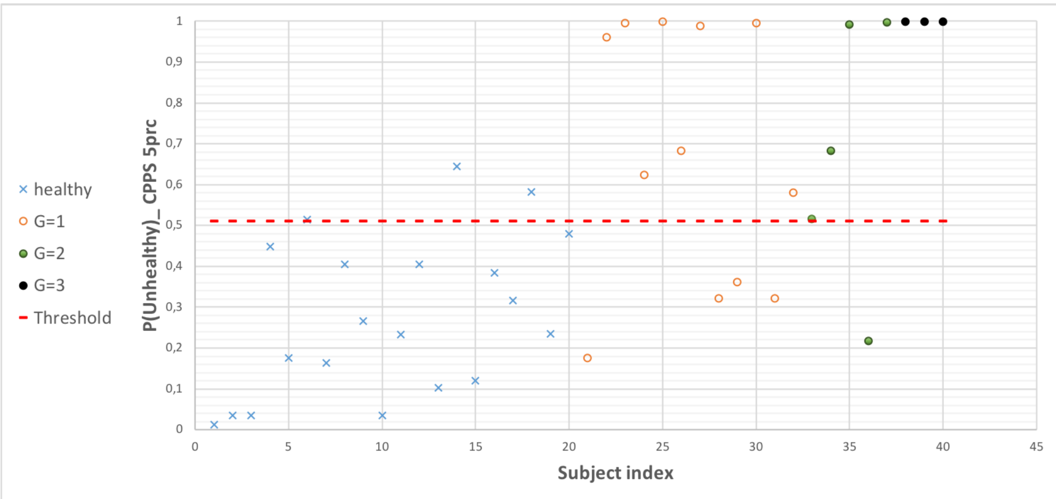

Figure 4.8: Fitted values of the logistic regression model for CPPS 5prc. Cross points indicate the healthy group; circle points represent the unhealthy group, where different colours and sizes represent subjects with different overall grade of dysphonia. The red line represents the threshold value of 0.51, which best separates healthy and unhealthy subjects.

The figure 4.9 shows the values provided by the logistic model, with the PPQ like independent variable, for each subject. In this case the classification improves compared to before, in fact all patients with G=3 are close to score 1, patients with G=2 are above the threshold and only three subjects with G=1 are incorrectly classified. Moreover, most of the healthy patients (cross points) have lower scores close to 0, with only 2 patients classified incorrectly and one on the threshold. The figure also shows the best classification threshold in terms of P(Unhealthy)=0.36.

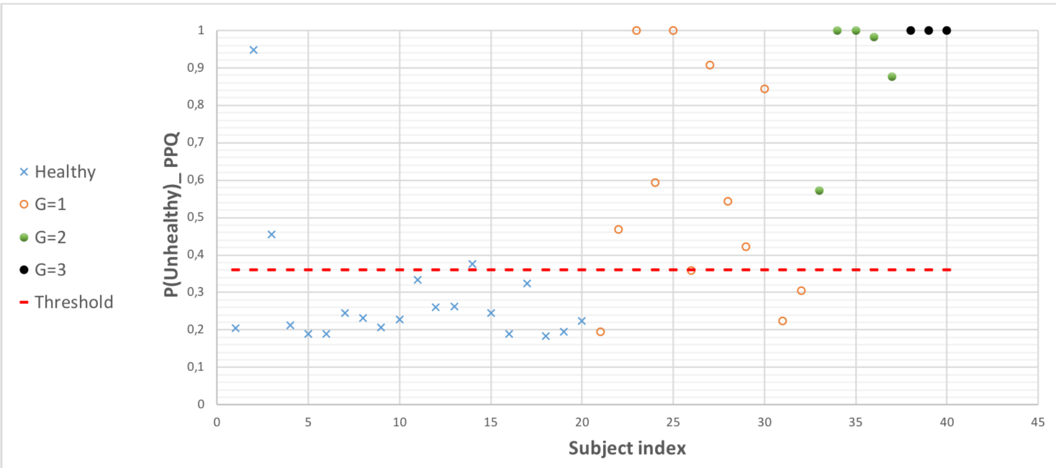

Figure 4.9: Fitted values of the logistic regression model for PPQ. Cross points indicate the healthy group; circle points represent the unhealthy group, where different colours and sizes represent subjects with different overall grade of dysponia. The red line represents the threshold value of 0.36, which best separates healthy and unhealthy subjects.

The figure 4.10 shows the values provided by the logistic model, with the CPPS5prc + PPQ like independent variable, for each subject. In this case the classification power increases. In fact all patients with G=3 and G=2 are close to score 1 and only two subjects with G=1 are incorrectly classified. Moreover, only 1 healthy patient is classified incorrectly. The figure also shows the best classification threshold in terms of P(Unhealthy)=0.41.

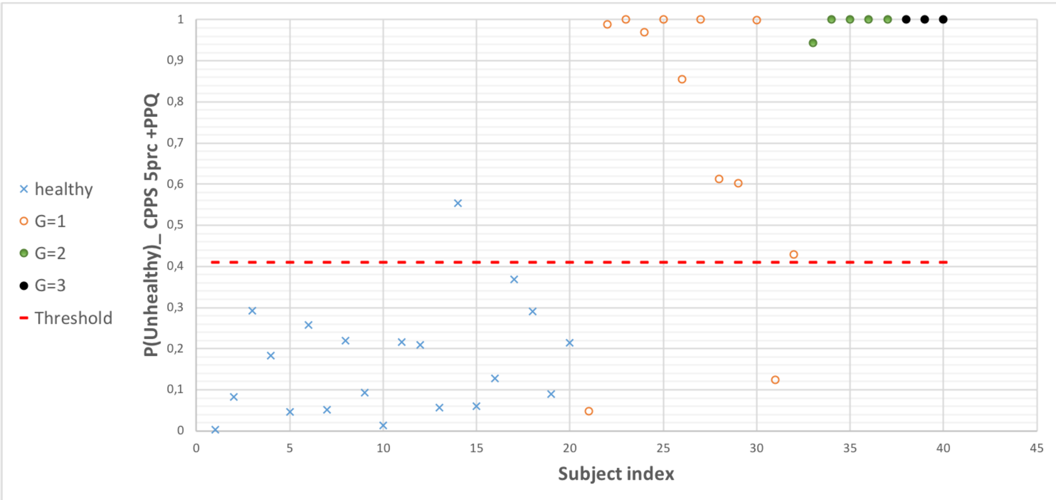

Figure 4.10: Fitted values of the logistic regression model for CPPS5prc+PPQ. Cross points indicate the healthy group; circle points represent the unhealthy group, where different colours and sizes represent subjects with different overall grade of dysphonia. The red line represents the threshold value of 0.41, which best separates healthy and unhealthy subjects.

Eventually, in figure 4.11 is shown again the fitted model of fig.4.10 but with triangles points in addition. These points indicates the patients with paralysis which are all very close to one for exception of one patient that is located on the threshold. This result is promising on a future discrimination based on the disease. Firstly, this type of study must be supported by a greater number of patients.

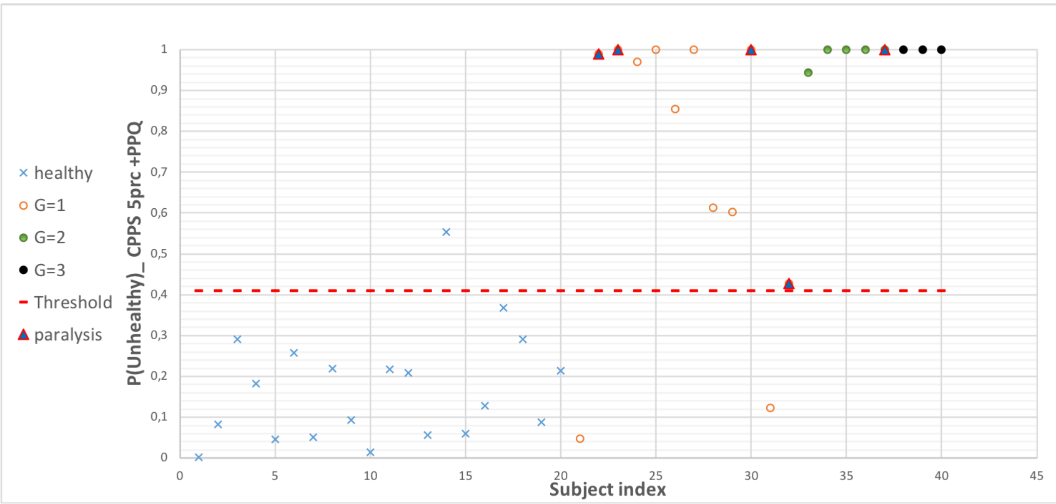

Figure 4.11: Fitted values of the logistic regression model for CPPS5prc+PPQ. The triangles points represents subjects affected by vocal paralysis.

Chapter 5

Conclusions

The main goal of this thesis consisted in providing an objective support in vocal folds surgery rehabilitation through the monitoring of the vocal signal and the estimation of voice-related parameters. A group of 8 patients affected by different vocal diseases were subjected to three monitorings at three different times: the first just before the operation, the second and the third ones 1 month and 2 months after the operation, respectively.

The patients were asked to vocalize the vowel /a/ three times, read aloud a passage, to speak about a free topic and to fill in a questionnaire. During the tasks the vocal signal and the neck-skin vibration were simultaneously acquired by means of a microphone in air and two contact microphones, a piezoelectric device and a Electret Condenser Microphone (ECM). During the processing phase, Cepstral Peak Prominence Smoothed (CPPS) descriptive statistics were estimated, both for vowels /a/ and continuous speech and frequency perturbation parameters, amplitude perturbation parameters, and Harmonic-to-Noise Ratio were estimated only for vowel /a/. The identified subset includes two parameters related to frequency perturbation (local Jitter - Jitt - and Pitch Period Perturbation Quotient - PPQ), two parameters related to amplitude perturbation (local Shimmer - Shim - and Peak Amplitude Variation - vAm), the Harmonic-to-Noise Ratio (HNR), and the 5^{th} percentile and the standard deviation of the CPPS distributions. The same subset of parameters have been evaluated for vowels /a/ acquired with the three microphones. For continuous speech, the investigated parameters was the 95^{th} percentile (for microphone in air), standard deviation (for the three microphones) and mean (for the contact microphones) of CPPS distributions. This choice is based on results that exists in literature. As a preliminary validation check, the parameters evaluated with a specifically developed Matlab script were compared to the commercial software environment Multi-Dimensional Voice Program (MDVP). This validation involved both pathological and healthy subjects and was performed for vowels /a/ acquired with the microphone in air. The results showed an high correlation for Shim and vAm parameters, with R^2 values of 0,97 and 0,99, and a lower correlation for Jitt and PPQ ones, with R^2 of 0,52 and 0,37, respectively. Consequently to these results, the thresholds for discrimination between unhealthy and healthy voice was extracted through the linear regression for each parameter, starting from the thresholds provided by MDVP manual. The values obtained are all lower than the one provided by the MDVP manual and in particular for Jitt the value was 0,78 % and for Shim was 3,63 %. Then, the pre-surgery and two months post-surgery results were compared for each speech material. In particular, for vowel /a/, the Mann-Whitney U-test has identified the PPQ as the parameter with the highest discrimination power between pre-surgery and post-surgery voice, with p-value = 0.001, both for microphone in air and contact microphone. Positive results, with p-values < 0.05 were obtained also for Jitt, Shim and HNR only for microphone in air.

For continuous speech in most cases the CPPS distributions before operation have shown higher occurrences at lower values, since for pathological voice most speech becomes noisy sounds. The CPPS distribution of the same subjects, after operation, showed a longer left tail and $CPPS_{\text{mean}}$ values higher

than those before operation. The direct observation of distributions could become a first immediate classification of patients health. Another important aspect emerged is that the CPPS distribution in most cases show overlapping results for reading and free speech tasks. Such a reproducible characteristic has been labeled in literature as “ CPPS vocal print”.
The correlation coefficients between piezoelectric microphone and microphone in air showed high values especially for CPPS parameters in reading and free speech tasks. Also the ECM microphone showed a high correlation with the microphone in air especially for reading and free speech tasks, but certainly this evidence need to be supported by an increase of the data-set. The positive results obtained with the piezoelectric contact microphone are forward-looking to extend this study to everyday life monitorings.
The PAPV mean scores obtained after operation were lower than those obtained before especially in the emotional effects section, so the patients had an overall better perception of their vocal conditions.
In the last part of this thesis a statistical analysis for discrimination of healthy and pathological voices in sustained vowel /a/ was carried out. The study involved 40 subjects (20 unhealthy and 20 healthy) which are part of already collected data-base. For each subject the CPPS parameters related to the first vowel /a/ were collected and, in addition, the Jitt, PPQ, Shim and vAm parameters were evaluated. A single-variable regression model was firstly performed with one parameter at a time. The logistic regression models that showed the best capability in discriminating healthy from pathological voice were the one which uses the parameter PPQ, with AUC = 0,89 and Accuracy = 85 %, and the one which uses $CPPS_{5\mathrm{prc}}$, with AUC = 0,87 and Accuracy = 84 %. Then, a double-variable logistic regression model was performed. The $CPPS_{5\mathrm{prc}}$ and PPQ, that individually showed the best performance were coupled and showed a better capability in discriminating healthy and pathological

voices as dependent variables of the double-variable logistic model. The results showed AUC equal to 0,93 and an accuracy of 86%. The unhealthy probability values of subjects affected by paralysis of the vocal folds were analyzed and showed values very close to one, except for one that anyway showed a probability value over the threshold. This result is promising on a future discrimination based on the type of disease that must be supported by a greater number of patients. Eventually, the logistic regression was carried out with three parameters: $CPPS_{\mathrm{5prc}}$, PPQ and Shim. One should note that a more complex model has not provided better results. Values of 0,93 for AUC and 86% for the accuracy were obtained again, respectively. Moreover, this study can be extended to the analysis of continuous speech with contact microphones that in previous studies has not provided good results using single-variable logistic regression.

Appendix A

Italian phonetically balanced passage

Avevo un bulldog che si chiamava Bulka. Era tutto nero salvo una macchia bianca allestremita delle zampe anteriori. Nei cani di questa razza, la mandibola e sempre prominent,cosi 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 dellavversario, 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 afferro tra i denti un orecchio. L‘orso cercava di colpirlo con una zampa, scuoteva la testa, ma non se ne poteva sbarazzare: fini per rovesciare il testone in terra per schiacciarvi il cane. Su questultimo, pero, perche 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 gettó su di me, mi lecco la mano e poi, la lingua ciondoloni, si stese allombra sotto la vettura. Seppi piu tardi che aveva rotto un vetro per seguirmi; era saltato dalla finestra: aveva percorso venti chilometri destate, sotto un sole bruciante.

Bibliography

[1] Brad H. Story. *The handbook of speech production*. John Wiley Sons Inc, 2015.

[2] W. De Colle. *VoceComputer, Analisi acustica digitale del segnale verbale (Il sistema CSL-MDVP)*. Omega Edizioni, 2001.

[3] Youri Maryn, Paul Corthals, Paul Van Cauwenberge, Nelson Roy, and Marc De Bodt. Toward improved ecological validity in the acoustic measurement of overall voice qual- ity: combining continuous speech and sustained vowels.*Journal of voice*, 24(5):540 555, 2010.

[4] Meike Brockmann, Michael J Drinnan, Claudio Storck, and Paul N Carding. Reliable jitter and shimmer measurements in voice clinics: the relevance of vowel, gender, vocal intensity, and fundamental frequency effects in a typical clinical task.*Journal of voice* , 25(1):4453, 2011.

[5] Joao Paulo Teixeira, Carla Oliveira, and Carla Lopes. Vocal acoustic analysis jitter, shimmer and hnr parameters. *Procedia Technology*, 9:11121122, 2013.

[6] P. Gomez-Vilda, R. Fernandez-Baillo, A. Nieto, F. Diaz, F.J. Fernandez-Camacho, V. Rodellar, A. Alvarez, and R. Martnez. Evaluation of Voice Pathology Based on the Estimation of Vocal Fold Biomechanical Parameters. *Journal of voice*, Vol. 21, No. 4, pp. 450476, 2006.

[7] M. Nicastri, G. Chiarella, L.V. Gallo, M. Catalano, E. Cassandro. Multidimensional Voice Program (MDVP) and amplitude variation parameters in euphonic adult subjects. Normative study. Department of Experimental and Clinical Medicine G. Salvatore, Audiology Unit, Magna Graecia University, Catanzaro, Italy, 2006.

[8] Yolanda D Heman-Ackah, Deirdre D Michael, and George S Goding. The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of voice*, 16(1):2027, 2002.

[9] James Hillenbrand, Ronald A Cleveland, and Robert L Erickson. Acoustic corre- lates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4):769778, 1994.

[10] James Hillenbrand and Robert A Houde. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39(2):311321, 1996.

[11] Yolanda D Heman-Ackah, Deirdre D Michael, and George S Goding. The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, 16(1):2027, 2002.

[12] Youri Maryn and David Weenink. Objective dysphonia measures in the program praat: smoothed cepstral peak prominence and acoustic voice quality index. *Journal of Voice*, 29(1):3543, 2015.

[13] Antonella Castellana, Alessio Carullo, Simone Corbellini, Arianna Astolfi, M Spadola Bisetti, and J Colombini. Cepstral peak prominence smoothed distribution as discriminator of vocal health in sustained vowel. *Instrumentation and Measurement Technology Conference (I2MTC)*, 2017 IEEE International, pages 16. IEEE, 2017.

[14] Antonella Castellana , Alessio Carullo, Simone Corbellini, and Arianna Astolfi. Discriminating Pathological Voice From Healthy Voice Using Cepstral Peak Prominence Smoothed Distribution in Sustained Vowel. *IEEE Transaction on Instrumentation and Mesaurement*. IEEE, 2018.

[15] Antonella Castellana , Alessio Carullo, Arianna Astolfi, Massimo Spadola Bisetti, and Jacopo Colombini. Vocal health assessment by means of Cepstral Peak Prominence Smoothed distribution in continuous speech. In *IEEE Instrumentation and Measurement Society*. IEEE, 2018.

[16] Gaetano Fava, Nico Paolo Paolillo, Gisele Oliveira, and Mara Behlau. Cross-cultural adaptation of the italian version of the voice activity participation profile. In CoDAS, volume 26, pages 252255. SciELO Brasil, 2014.

[17] Multi-Dimensional Voice Program (MDVP) Model 5105, KayPENTAX Lincoln Park, NJ 07035-1488 USA

[18] Paul Boersma. Accurate short-term analysis of the fundamental frequency and the harmonic to noise ratio of a sempled sound. IFA Proceedings 17, 1993.

[19] H. Oguz, M. A. Kilic, M. Safak. Comparison of results in two acoustic analysis programs: Praat and MDVP. *Turk J Med Sci*, 2010.

[20] Ferrero F. E., Lanni R.,De Colle W. Primi risultati di uno studio per la validazione del sistema MDVP come strumento per una caratterizzazione multiparametrica della voce. Acta Phon. Lat., vol. 17, 161-180, 1995.

[21] A. Castellana, A. Carullo, A. Astolfi. Dysphonia Recognition in Reading and Free-Speech through Cepstral Peak Promincence Smoothed Estimated by Microphone in Air and Contact Microphone. Submitted for publication.