POLITECNICO DI TORINO

Master's Degree in Biomedical Engineering

Master's Degree Thesis

Objective Evaluation of Speech Therapies for Multiple Sclerosis Patients

An approach based on vocal indexes obtained with different software applications

Supervisors

Prof. Alessio CARULLO

Prof. Alberto VALLAN

Candidate

Tiziana BASILEO

October 2024

Abstract

Multiple Sclerosis (MS) is a chronic demyelinating autoimmune disease that affects the central nervous system (CNS) causing a wide spectrum of signs and symptoms: motor symptoms such as spasticity and weakness, fatigue, dysphonia, speech or vocal disorders.
Hypophonia, which is characterized by a reduction in the intensity of the voice, has a particular relevance among all the signs of the disease which demands for vocal analysis and therapies that allow to reduce social life limitations of MS patients.
One of the most advanced hospital facilities that is involved in the MS research is the therapy and rehabilitation department of Don Carlo Gnocchi Foundation in Milan.
This foundation deals with speech therapy and among these the Lee Silverman Voice Treatment (LSVT) therapy acquires particular relevance. This therapy is administred in an innovative way in telerehabilitation to ten MS subjects, allowing a comparison with the classic LSVT therapy administred in clinic to other ten patients.
Therefore, the aim of the analysis conducted in this thesis is to evaluate the non inferiority of the telerehabilitation therapy compared to the in-clinic one, opening the possibility of using a therapy without problems of accessibility, waiting times and costs.
This analysis is accomplished analyzing the speech material composed of three repetitions of a sustained vowel /a/, a free speech and a reading recording. These recordings are acquired with a vocal recorder equipped with an in-air microphone and with a contact microphone (Vocal Holter, VH) before the therapy (T0), after the therapy (T1) and after three months from the end of the therapy (T2). Since T2 recordings are not available for a sufficient number of patients, the analysis is focused entirely on sounds at T0 and T1.
The evaluation is conducted extracting from these recordings the most significant vocal parameters that are Jitter, Shimmer, HNR, CPPS and SPL and computing the vocal indexes Acoustic Voice Quality Index (AVQI) and Warning Score (WS). This is performed using software commonly employed in the vocal analysis that are Praat, VOXplot and Matlab, analyzing their pre-processing, parameters extraction criteria and output.
The analysis reveals similarities between Praat and VOXplot in terms of AVQImean $\pm$standard deviation, respectively returning values of 6.0 ± 0.29 and 6.0 ± 0.30 at T0. Also Matlab provides a similar output, resulting in an AVQI of 5.5 ± 0.26. These AVQI values are compared to those obtained at T1, that are 5.2 ± 0.29, 5.1 ± 0.29 and 4.9 ± 0.27 respectively for Praat, VOXplot and Matlab: this confirms

a slight improvement of the disease after the therapy, considering that a lower AVQI corresponds to a healthier subject.
In addition, the obtained results confirm the effectiveness and the reliability of the telerehabilitation therapy compared to the in-clinic one. Considering as example the delta AVQI (difference between T0 and T1) in Praat, the value of 0.5±0.19 for the telerehabilitation group is similar to the one obtained for the in-clinic group and equal to 0.9±0.20.
The same analysis conducted for the AVQI is done in terms of WS, confirming also in this case the improvement of the disease for some patients as well as the effectiveness of the telerehabilitation. Finally, a comparison in terms of WS is conducted between the two types of microphones used for the acquisitions of the recordings.

Acknowledgements

I would like to thank Prof. Alessio Carullo for his commitment and passion as thesis advisor, and the Don Gnocchi Foundation Hospital in Milan, which made available the voice recordings of multiple sclerosis patients.

A great recognition goes to my family and my friends for their continuous presence, and to Luigi for his endless support and encouragement.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter introduces a general overwiew of the phonatory system and its mechanism of voice production, underlyning the vocal pathologies related to it.
In particular, between all the pathologies, the focus is on the Multiple Sclerosis (MS) disease with its vocal symptoms and therapies.
In the therapies context, the Lee Silverman Voice Treatment LOUD (LSVT-LOUD) therapy is relevant and it is described bacause it is the one used to treat hypophonia and vocal fatigue in MS patients.

1.1 Phonatory System and Voice Production

The voice production is a complex mechanism that involves the phonatory system, a system composed by various elements:

- *Lungs*: they provide the air flow necessary for the production of the voice.
- *Vocal folds*: they are located in the larynx and they have the function of producing the voice.
- *Vocal tract*: it is a cavity responsible for the modulation and creation of the vocal sounds.

An overwiew of the vocal tract is shown in Figure 3.15.
The phonation cycle involves the alternation of an open phase and a closed one. Durign the open phase the glottis is open and during the closed one the glottis can be totally or partially closed.
The phonation process begins when the vocal folds come closer together creating a partial or total closure of the glottis. Later, when the lungs are contracted, a pressure is created under the glottis and the air is pushed out. When the pressure

becomes quite high exceeding a threshold, the vocal folds vibrate. These vibrations turn the airflow into a series of pulses. When this airflow becomes turbolent, it moves through the vocal tract, producing a sound [1].

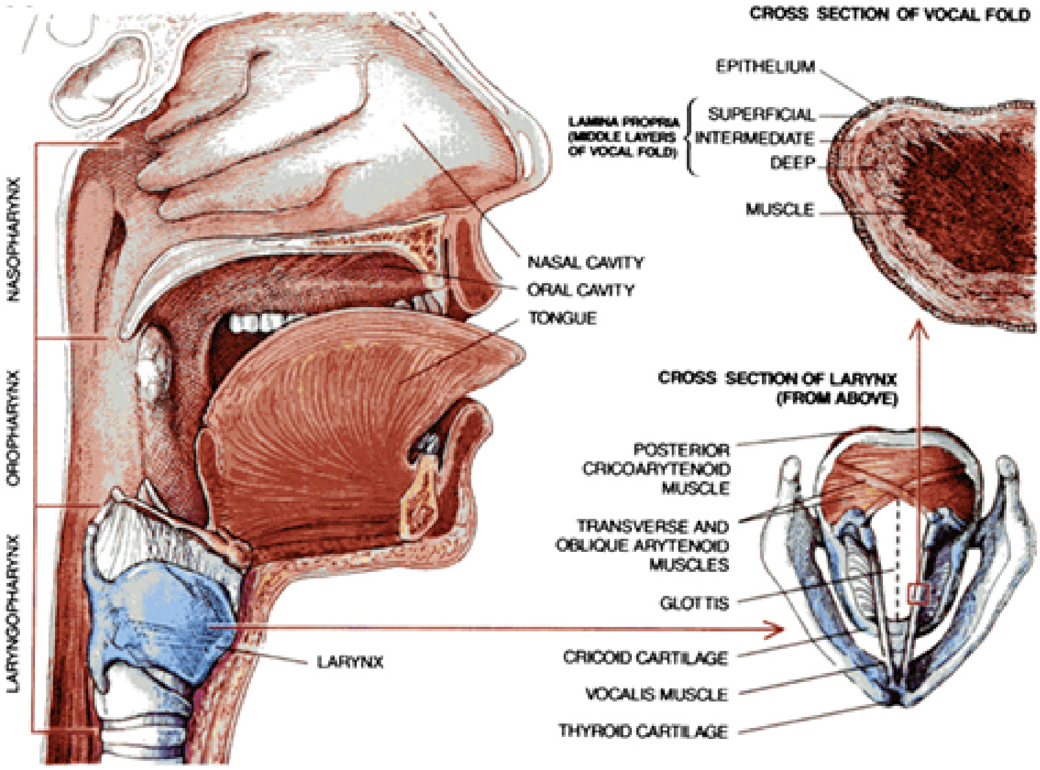

Figure 1.1: Overwiew of the vocal tract [2].

A subject can be affected by various pathologies involving the vocal tract. Of relevance in this thesis is the Multiple Sclerosis (MS), as subjects affected by the latter will be analysed.

1.2 Multiple Sclerosis

Multiple Sclerosis (MS) is a chronic illness that affects the central nervous system (CNS), causing infiammation and demyelination.
This disease is tough to recognize in the initial stages, due to very mild symptoms such as tingling, vision loss, weakness, imbalance, incoordination and numbness. It is diagnosed thourgh a series of tests such as magnetic resonance imaging (MRI),

cerebrospinal fluid analysis and evoked potentials, always considering the patient's clinical history [3].
The MS is more frequent in women and not in man (ratio of 2.5:1).
It is a pathology that progresses over time and it is why it is classified with different scales of severity: RRMS (Relapsing Remitting), PPMS (Primary Progressive), SPMS (Secondary Progressive) and PRMS (Progressive Relapsing).
The MS disease involves a wide spectrum of signs and symptoms, as they depend on the area involved by the demyelination of the neurons of the CNS. The most common symptoms are [4]:

- *Motor symptoms*: weakness, spasticity, loss of balance and tremor.
- *Sensory symptoms*: numbness, tingling, pain and itching.
- *Visual symptoms*: optic neuritis and diplopia.
- *Cognitive symptoms*: depression and difficulty in learning.
- *Vocal symptoms*: hypophonia, dysphonia, speech or vocal disorders.

Among all these symptoms, vocal ones acquire particular relevance: it is estimated that 62% of sufferers are affected by them.
Specifically, they may include weakness of the respiratory system, partial glottal closure, posterior glottal gap, hypophonia, vocal asthenia, harshness and breathiness.
Hypophonia, characterized by a reduction in the intensity of the voice, has a particular relevance among all the vocal signs of the disease, as it is considered one of the most significant issues by 16% of individuals and because it appears in the initial phases of the MS evolution [5].
This symptom can be treated with the Lee Silverman Voice Treatment LOUD (LSVT-LOUD) therapy, described in the next section, in order to improve the quality of life of the MS patients.

1.3 Lee Silverman Voice Treatment LOUD

In the field of the vocal therapy, the Lee Silverman Voice Treatment LOUD (LSVT-LOUD) is an effective treatment for hypophonia and vocal fatigue in MS patients.
The LSVT-LOUD is a therapy based on the principle of neuroplasticity and on the one related to the motor learning, because they are the two principles at the basis of neurological diseases. Theferore, this therapy is commonly used in subjects affected by Parkinson's disease, stroke, brain injury, cerebral palsy and MS [6].

The main purpose of the LSVT-LOUD therapy is the enhancement of the vocal intensity, reached by training the intensity of the voice. In particular, it aims to improve the activity of the muscles involved in the production of the voice.
This treatment involves high-effort tasks (free speech exercises) combined with simple and repetitive ones (repetition of a sustained vowel /a/ and reading of sentences) [7].
The LSVT-LOUD therapy is normally administred to the subjects in clinic, but currently the administration in telerehabilitation is being tested, as discussed in the next section.

1.4 Telerehabilitation LSVT-LOUD

The use of the LSVT-LOUD treatment is relevant to improve the loudness of the voice and to reduce the vocal fatigue. Unfortunately, only 2% of MS patients undergo speech therapy, due to various difficulties such as costs, waiting times and overwhelming demand on health services. Therefore, the telerehabilitation system is developed to overcome these difficulties.
The LSVT-LOUD, normally administred in clinic, is in this case (Tele-LSVT-LOUD) adimistred remotely. The patient just has to access to an online platform at home and follow the same 4-weeks therapy [6].
There is currently no evidence of the validity of this method of administering the treatment. Therefore, in this context, the aim of this thesis will be to assess the non inferiority of the telerehabilitation treatment compared to the in-clinic one.

Chapter 2

Materials and Methods

This project of thesis is carried out in collaboration with the speech therapy and rehabilitation department of Don Gnocchi Hospital in Milan that provided the dataset that is used for the analyses conducted in this study.
Don Gnocchi hospital was founded in 1945 thanks to the commitment of Don Carlo Gnocchi.
Today it is one of the largest private hospital in the national territory and it counts twenty-five residential structures and twenty-seven clinics diffused across nine regions.
The main purpose of this structure is to take care of people who suffer considering that their illness comes after their person and it is done in collaboration with healtcare figures, volunteers and caregivers.
Don Gnocchi foundation operates in different fields such as rehabilitation, imaging and instrumental diagnostics.
This structure offers services for the developmental age and adolescents, people with acquired brain injures, the elderly, the disabled and the terminally ill.
In particular, the foundation deals with speech therapy and among these the Lee Silverman Voice Treatment (LSVT) - LOUD therapy acquires particular relevance.
This therapy is used to treat the hypophonia that characterize Multiple Sclerosis (MS) patients, which are one of all the categories of subjects treated by the foundation and this is the category of patients analyzed in this project of thesis.
The LSVT-LOUD therapy is administred in an innovative way in telerehabilitation (Tele-LSVT-LOUD) to ten MS subjects, allowing a comparison with the classic LSVT therapy administred in clinic to other ten patients. So the foundation provided a dataset of totally twenty patients with diagnosed MS divided in a random way in the in-clinic group and in the telerehabilitation group for the therapy administration.
Therefore, the main purpose of the analysis conducted in this thesis is to evaluate

the non inferiority of the telerehabilitation therapy compared to the in-clinic one, opening the possibility of using a therapy without problems of accessibility, waiting times and costs.
This aim is reached analyzing the speech material composed of three recordings of a sustained vowels /a/, a monologue recording and reading passage recording, acquired with both a vocal recorder, equipped with an in-air microphone, and with a contact microphone (Vocal Holter, VH).
The study is conducted in terms of vocal parameters normally extracted in literature from the monologue and the sustained vowel recordings, and in terms of vocal indexes that are Acoustic Voice Quality Index (AVQI) and Warning Score (WS).
These parameters are extracted using software apllications commonly employed in the vocal analysis that are Praat and VOXplot, and also using Matlab. This is done in order to aldo evaluate their pre-processing, parameters extraction criteria and output.
Therefore, a comparison between the different software is conducted in terms of AVQI mean and vocal parameters, extracted from all the patients with the different software applications.
Then, the ΔAVQI and the Δ values of some vocal parameters, avereged distinctly for the two groups of subjects, are used to evaluate the non-inferiority of the telerehabilitation therapy compared to the in-clinic one. Also the other vocal index, the WS, is used to reach this last aim.
The WS is finally employed to compare the two types of acquisition systems, allowing to evaluate and to assess the reliability of the contact microphone, compared to the in-air one, in the extraction of the vocal parameters and so, of the considered vocal index.

2.1 Dataset

The dataset, provided by Don Carlo Gnocchi foundation, includes twenty patients affected by Multiple Sclerosis (MS) of different severity. The dataset is composed of patients of different genders and ages.
These subjects are treated using the Loud Silverman Voice Treatment - Loud (LSVT-Loud) therapy, which is functional to alleviate the symptom of hypophonia that is tipical of this group of ill, with the aim to increase the voice intensity in MS patients.
A distinction between the patients is made on the basis of the therapy administration that has been used to treat the MS. The LSVT-Loud method is administred in clinic to ten of these patients (LSVT-Loud) and to the other group of ten patients the same therapy is administred in telerehabilitation (Tele-LSVT-Loud). In this way a control group (in-clinic subjects) and an experimental group (telerehabilitation

subjects) are created. The choice of subjects who must belong to the experimental or to the control group is random.
The patients in telerehabilitation follow a 4-week therapy accessing a platform at home and the patients in presence follow the same therapy but delivered in clinic, as it is done conventionally [6].
A diagram of the work plan is shown in Fig. 2.1.

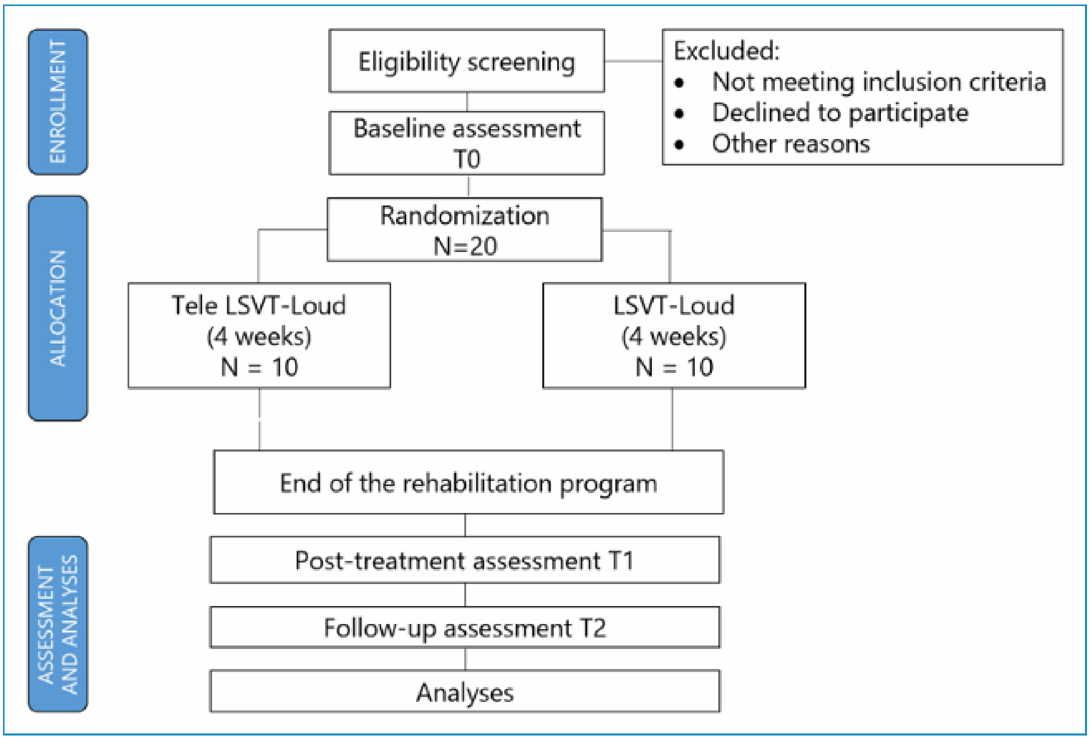

Figure 2.1: Trial work plan [6].

Table 2.1 shows the identification code (ID), the gender (M=male, F=female) and the treatment followed (Telerehabilitation or In-clinic).
All the acquisitions are done with both an in-air microphone and a contact microphone (Vocal Holter, VH).

ID	Gender	Treatment
01	F	In-clinic
02	F	Telerehabilitation
03	F	In-clinic
04	M	Telerehabilitation
05	F	In-clinic
06	M	In-clinic
07	M	Telerehabilitation
08	F	In-clinic
09	F	Telerehabilitation
10	M	Telerehabilitation
11	F	In-clinic
12	F	Telerehabilitation
13	F	In-clinic
14	F	Telerehabilitation
15	F	In-clinic
16	F	Telerehabilitation
17	F	Telerehabilitation
19	M	In-clinic
20	M	In-clinic
21	F	Telerehabilitation

Table 2.1: Dataset of multiple sclerosis patients.

For each patient various recordings are available:

- *Three repetitions of a sustained vowel /a/.*

  The duration of the maintaining of each repetition of vowel is different between the same patient and between different patients, but it is always longer than two seconds.

- *A monologue recording.*

  All the subjects had to respond to the same questions, so the duration of this recording is different between each patient, but it lasts at least one minute.

- *A reading passage.*

  Each patient had to read a phonetically balanced piece called 'Notturno':

  "Notturno. Vi è un profondo silenzio nel buio della notte. Vicino al pozzo, nella cui acqua si specchiano la luna ed una scia di stelle, la magnolia stende i suoi rami, cespugli di rose olezzano nell'aria. Il temporale è cessato e la pioggia, ormai, non cade più. Solo le rane gracidano nei fossi oltre quel prato."

This set of recordings is acquired for each subjects at three different times:

- *T0*

  Acquisition made after recognition of the disease and before starting the treatment. It allows to have a baseline idea of the severity of the disease of the patient and to evaluate future improvements or worsening.

- *T1*

  Acquisition made after the treatment, whether it is administred in-clinic or in telerehabilitation.

- *T2*

  Acquisition made after three months from the treatment (after the time T1). It is useful to evaluate the maintanance of the possible improvements after the therapy. In this thesis, this recording will not be used because it isn't available for all the patients but only for eight of them (four for the control group and four for the experimental group).

2.2 Acquisition Systems

The dataset, made available by the Don Gnocchi foundation, includes for each subject a set of recordings (three sustained vowel, a reading passage and a monologue recording) that is acquired with both a vocal recorder, equipped with an in-air microphone, and with a contact microphone (Vocal Holter, VH).

2.2.1 Microphone In Air

The microphone in air used to obtain the recordings is an omni-directional microphone and it is included into a vocal recorder.
It is placed at a distance of about 30cm from the lips of the patient and it captures the pressure of the air coming out of the mouth. Therefore, the signal obtained undergoes a filtering effect from the oral cavity and it is influenced by the acquisition environment.
The microphone in air uses a vocal recorder that employs a sample rate of 44.1 KSa/s and 16 bit of resolution.
The output files from the recorder are in .wav format and they includes both moments of silence and speech in order to give the patient the freedom to record when they are ready.
The recordings are performed in an environment that is as quiet as possible ,to avoid the presence of superimposed noises that could alter the subsequent pre-processing of the signal and the extraction of the parameters of interest.

2.2.2 Contact Microphone

The contact microphone (Vocal Holter ,VH) is a type of microphone that has been developed at Politecnico of Turin.
It is a portable device that is composed of a contact microphone that is worn around the neck.
It has the function of detecting the vibrations created by the vocal cords during the emission of sentences (free speech or reading) or sustained vowels (commonly vowel /a/).
It allows to directly have in output parameters of interest for the acoustic and vocal analysis and it does not return an intelligible signal.
For these reasons, it is considered innovative and it differs from the in-air microphone, which is placed at a distance of a few centimeters from the mouth and returns recordings in .wav format.
The VH has the primary function of being a prevention tool allowing, as mentioned before, to extract parameters of interest in the vocal field. Its use as a diagnostic tool is also possible, combining the different parameters to obtain additional information on the pathological situation of a subject.
It uses a sample rate of 44.1Ksa/s and a 16-bit resolution [8].
In Fig. 2.2 the VH device is shown.

Figure 2.2: Example of a contact microphone [8].

The kit, that is shown in Fig. 2.3, is composed of various elements:

- A DAP unit (Data Acquisition and Processing) which is composed of a microphone and a spacer. The size of this unit is about 9x8x4cm.
- Contact microphone. The model is hx-505-1-1.
- Power adapter and cable
- Instruction manual

Figure 2.3: Elements of the vocal holter device [8].

Before using the device, it is necessary to connet it to the DAP unit and to wear it around the neck. Then, the DAP unit has to be switch on. Finally, the web interface have to be opened on the PC and the desidered operations can be selected. An example of web interface which is composed of various sections, is shown in Fig. 2.4.
Before the acquisition of the recording, as displayed in the first section, a calibration of the air microphone and of the contact microphone is necessary.
Addittionally, the environmental and system informations are displayed, i.e. the temperature, the humidity and the battery charge.
The data and parameters have to be dowloaded from the PC [8].

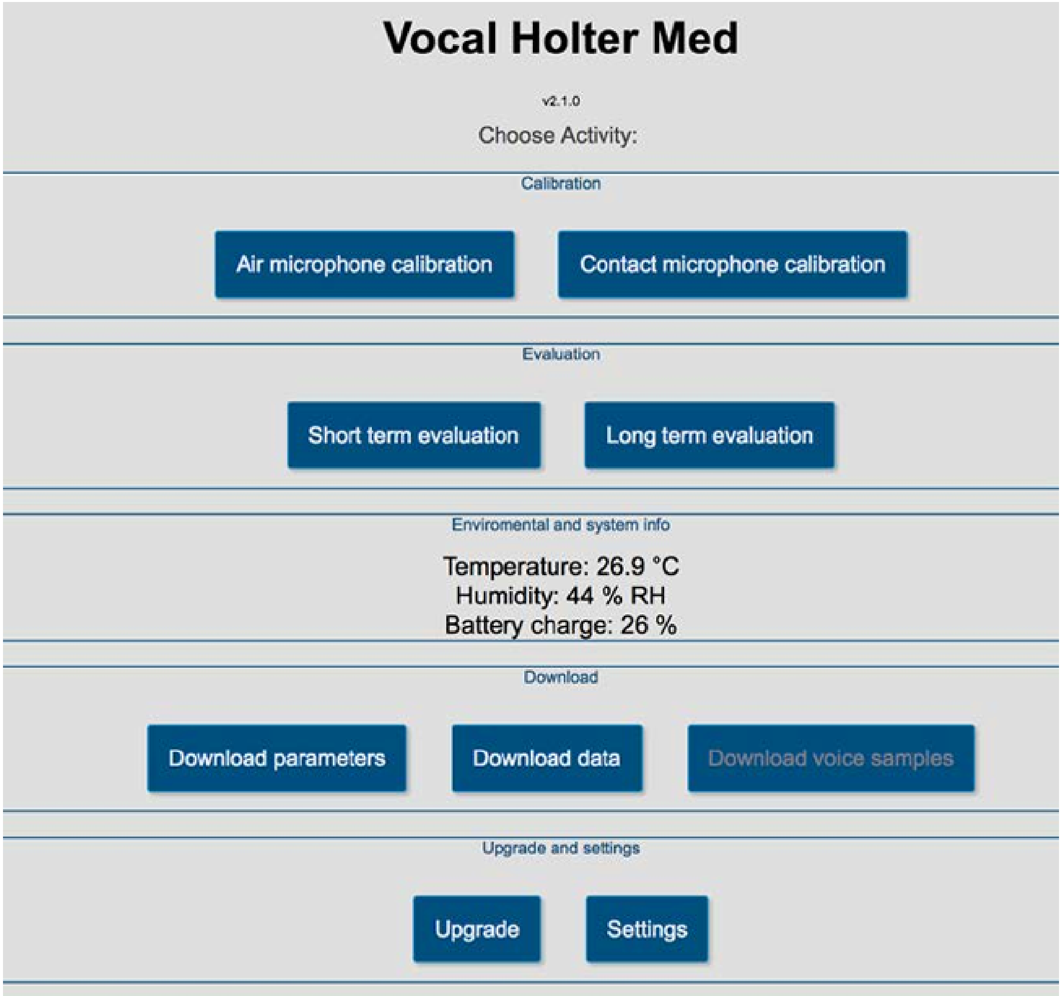

Figure 2.4: Example of the web interface of the vocal holter device [8].

The vocal parameters obtainable are:

- Local jitter (%)
- Local shimmer (%)
- Cepstral Peak Prominence Smoothed (CPPS) median (dB)
- CPPS std (dB)
- Fundamental frequency (Hz)
- Sound pressure level (SPL) in dB, at about 22cm from the mouth

In Fig. 2.5 an example of the output of the web interface at the end of a recording of a sustained vowel is shown.
The environmental parameters available are:

- Background Noise Level (dB)
- Air Temperature (°C)
- Air Relative Humidity

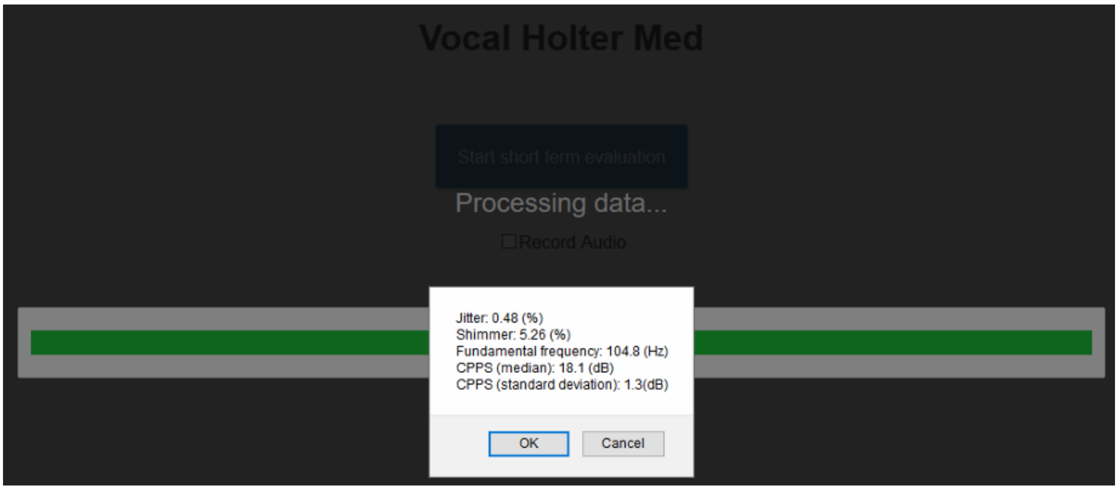

Figure 2.5: Example of the web interface of the vocal holter after the processing of a vowel [8].

The VH device allows to acquire both short-term and long-term recordings.
The short-term recordings are the ones that are used in this thesis project and on which parameters and vocal indexes are extracted. These are short acquisitions whose duration can vary from a few seconds, as in the case of the sustained vowels, to a few minutes, as in the case of the reading and the monologue recordings.
The long-term recordings have a much longer duration which can last a few hours. They are useful for the monitoring of the patient's voice during the day and to assess their fatigue. Infact, very often vocal problems in subjects with multiple sclerosis are not fully highlighted in a few minutes of speaking, but in a long-lasting one. These types of recordings are not used in this thesis project, but they could be used in future studies.

2.3 Software

The study conducted involves the use of various software applications commonly employed in literature in order to extract vocal parameters and vocal indexes (i.e. Acoustic Voice Quality Index and Warning Score).
The software applications commonly employed in the field of the vocal analysis are Praat and VOXplot; in addition, in this study, also Matlab is used.
This is finalized to evaluate the pre-processing, the parameters extraction criteria and the results of these software, making possible a comparison between them.
Matlab is employed in the version R2022b. Praat and VOXplot are described respectively in the subsections 2.3.1 and 2.3.2.

2.3.1 Praat

Praat is a software developed by Paul Boersma and David Weenink to the University of Amsterdam [9] for the audio analysis of voice and speech. It is relevant for the investigation, synthesis, and manipulation of speech, and it is now commonly used on a daily basis by phonologist and phoneticians.
It is available for free for all the main computer platforms (Macintosh, Windows, Linux, Raspberry Pi, Chromebook).
The main functionality of Praat enclose [10]:

- *Record and view a sound*

  It is possible with any input device. By selecting the apposite section 'View and edit' a screen opens: on the upper part the wave form of the sound is represented and on the bottom part the spectrogram (in shades of grey) is reported. On the latter, formant tracks (in red dots), an intensity curve (in green) and a pitch curve (in blue) are superimposed.

In Fig. 2.6 an example of the section for the visualization of the vocal sound is reported.

- *Generate a sound*

  It is possible to produce various types of audio signal, for example white noise and sine waves.

- *Label and segment a recording*

  It is possible to label and segment the speech recordings making transcriptions and annotations possible.

- *Develop and run scripts*

  It is possible to develop scripting codes in Praat language that allows to simulate the choices which can be selected in the menu by a user.

- *Draw a sound*

  It is possible to draw the sound and other data types (i.e. spectrograms, pith contours). In Fig. 2.7 an example of a Praat script and of a picture section is reported.

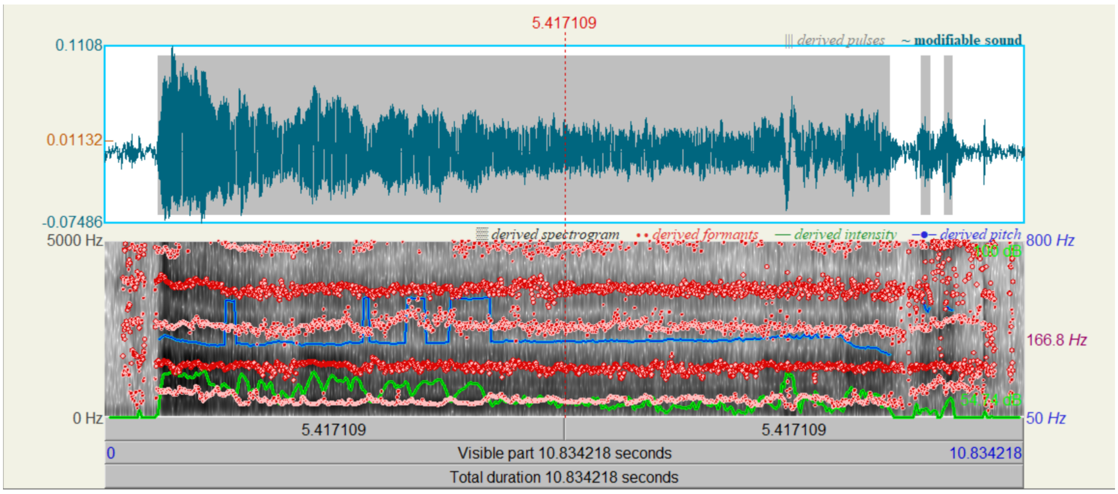

Figure 2.6: Example of a Praat section for the visualization of the wave form and of the spectogram of a sound.

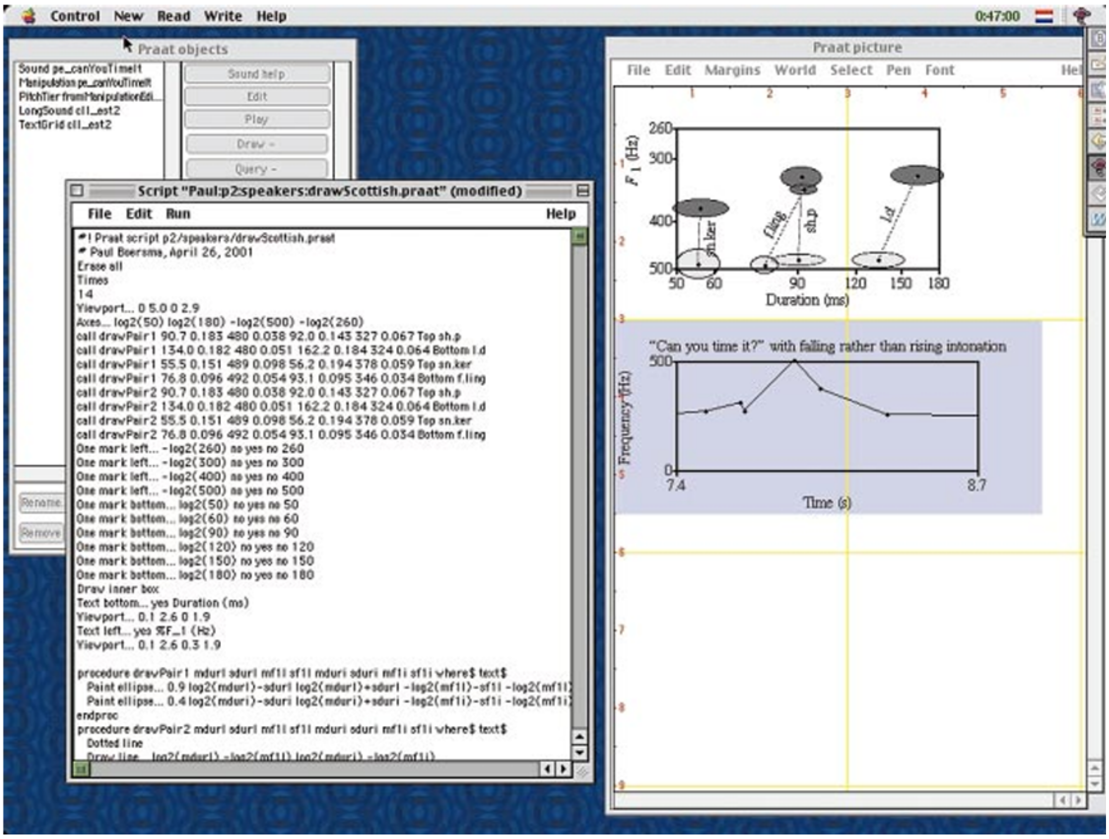

Figure 2.7: Example of a Praat script and a picture section [10].

2.3.2 VOXplot

VOXplot is an open source software designed for the acoustic voice analysis and based on reliable Praat algorithms.
This software presents a simple user interface, as shown in Fig. 2.8, with two different sections where is possible to record or upload a continous speech recording (a reading passage or a monologue) of any duration and/or a sustained vowel recording of three seconds.
A length of three seconds is necessary as the software allows the calculation of the Acoustic Voice Quality Index (AVQI), which requires this fixed duration of the sustained vowel.
If the length of the vowel file is longer than three seconds, it is possible to manually select this time [11].

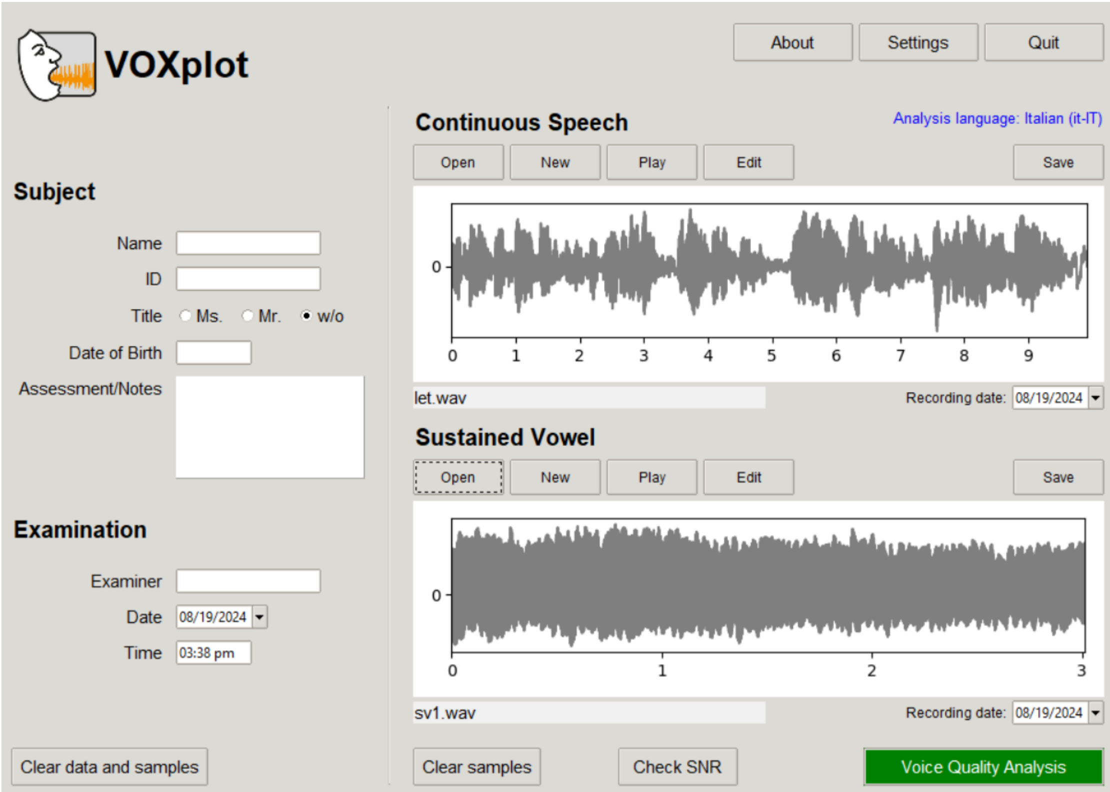

Figure 2.8: Example of VOXplot user interface.

VOXplot allows to calculate 19 acoustic parameters and two multidimensional indices: the Acoustic Voice Quality Index (AVQI) and the Acoustic Breathiness Index (ABI).
The columns of the extracted parameters are referred to the sustained vowel (SV), to the continous speech (CS) or to a combination of them (MX).
In particular, the results of the MX column are evaluated on a concatenation file of the SV and CS recordings. AVQI and ABI are computed only if both files SV and CS are available.
The results are displayed as numerical data and also in a graphical way with a norm-value circle [11].
Fig. 2.9 points out an example of VOXplot output.

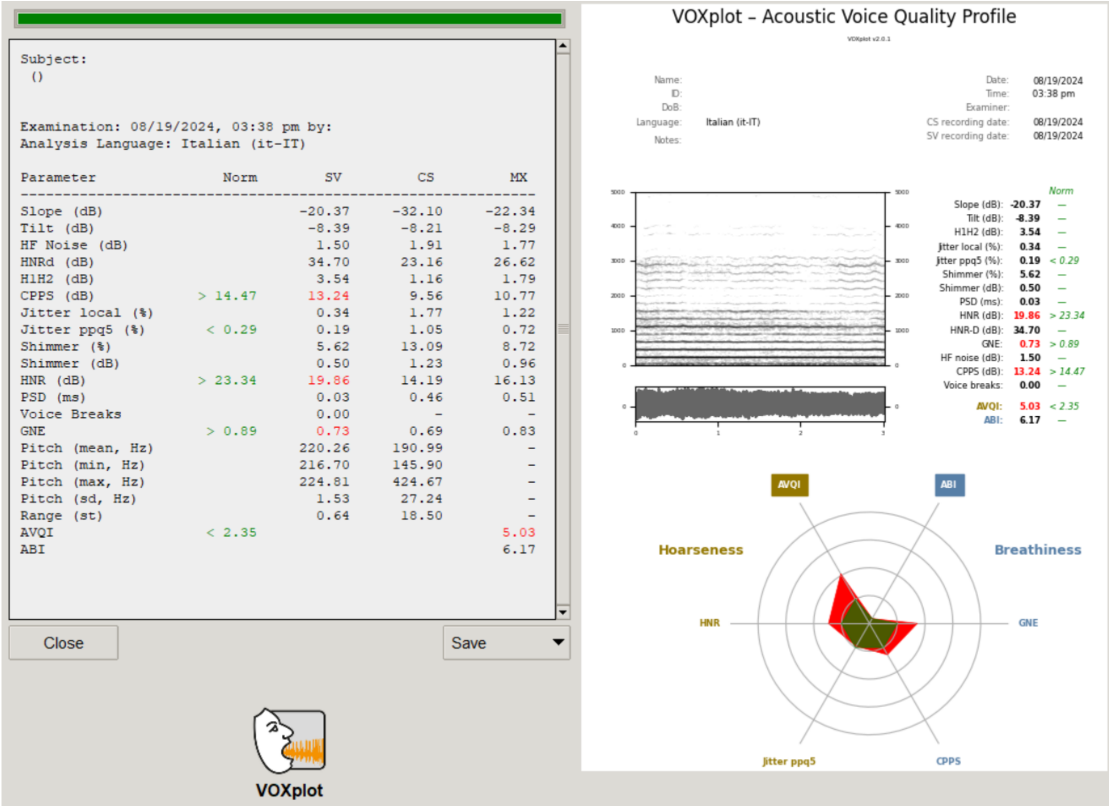

Figure 2.9: Example of VOXplot output.

2.4 Vocal Parameters

In the vocal analysis there are many extractable parameters. In particular, for the sustained vowels the parameters of interest in this analysis are Jitter, Shimmer, Harmonic-to-Noise Ratio (HNR), Fundamental Frequency (F0), Cepstal Peak Prominence Smoothed (CPPS).
Considering the monologue and the reading recordings, the parameters of interest are HNR, F0, CPPS and Sound Pressure Level (SPL).
Other parameters considered and which are used for the computation of the Acoustic Voice Quality Index (AVQI), described in the chapter 2.5, are the Spectal Slope and the Spectral Tilt.
Therefore, the parameters of interest are in the time domain, in the frequency domain and in the cepstral domain and they are described in the next subsections.

2.4.1 Time domain parameters

In this section the time domain vocal parameters explored in detail are:

- *Jitter Percent*

  It is a relative measurement of the cycle-to-cycle variation of the fundamental frequency. It is expressed in percentage [13].

  It can be computed as:

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

  where $T_0^{(i)}$ with i =1,2,...N is the extracted pitch period data and N is the number of extracted pitch periods.

  Being the Jitter Percent a relative measure, unlike Absolute Jitter, it is less dependent to the average fundamental frequency.

- *Shimmer in dB*

  It is a measure of the cycle-to-cycle variation of the peak-to-peak amplitude of the signal. It is expressed in decibels [13].

  It can be computed as:

$$Shimmer\ in\ dB = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20 \cdot \log\left(\frac{A^{(i+1)}}{A^{(i)}}\right)\right| \tag{2.2}$$

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

- *Shimmer Percent*

  It is a relative measurement of the cycle-to-cycle variation of the peak-to-peak amplitude of the signal. It is expressed in percentage [13].

  It can be computed as:

$$Shimmer\ Percent = 100\frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|\left(A^{(i)} - A^{(i+1)}\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \tag{2.3}$$

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

Both the Shimmer in dB and the Shimmer Percent are dependent and very sensitive to the amplitude and so, if there are errors in its estimation, the shimmer values can be overestimated or underestimated even significantly.

Both of them are relative measures of the amplitude variation but they adoperate a different unit for the result (dB or percentage).

- *Fundamental frequency F0*

  It can be defined, for a periodic signal, as the inverse of the period of the signal: $F_0 = 1/T_0$ [12]. It is expressed in Hz.

  The Average Fundamental Frequency ca be defined as the average value of all the frequency values extracted from the signal [13].

  It can be computed as:

$$F_0 = \frac{1}{N}\sum_{i=1}^{N} F_0^{(i)} \tag{2.4}$$

  where $F_0^{(i)}$ is the period-to-period fundamental frequency, $F_0^{(i)}$ with i =1,2,...N is the extracted pitch period data and N is the number of extracted pithc periods.

- *Root Mean Square (RMS)*

  It is computed as the square root of the arithmetic mean of the squares of the samples and it is expressed in "arbitrary units" (a.u.):

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{2.5}$$

  where x_i with i =1,2,...N is the signal data and N is the number of samples.

  The RMS introduction is necessary as this parameter is used in the preprocessing phase described in the chapter 2.6.

  Furthermore, it must be employed for the computation of the Sound Pressure Level (SPL).

- *Sound Pressure Level (SPL)*

  The SPL is used to measure the intensity of the sound. Because of this, it is useful to evaluate the intensity of the tone of the voice which is a parameter that can be affected by some pathologies.

  The sound travels through the air causing fluctuations in air pressure. This fluctuation is commonly known as sound pressure (p), measured in Pascals

(Pa). However, it is more typical to represent it using the logarithmic scale of SPL expressed in decibels (dB).

The SPL can be computed as:

$$SPL = 20 \cdot \log_{10}\left(\frac{p}{p_0}\right) \tag{2.6}$$

where p is the sound pressure computed as

$$p = 0.0027 + 6.0474 \cdot \mathrm{RMS}$$

and p_0 is the reference sound pressure which is equal to $20\mu Pa$. In this way, the value of SPL is equal to 0 dB when the sound pressure is equal to the reference sound pressure.

At this point the distance from the mouth must be considered. Indeed, if the distance increases, the sound disperses over a larger area and the pitch of the voice decreases. This concept is taken into account with the following formula:

$$SPL_d = SPL_{d_0} + 20 \cdot \log_{10}\left(\frac{d}{d_0}\right) \tag{2.7}$$

where SPL_d and SPL_{d_0} are the sound pressure level at the distance respectively d and d_0 from the mouth [14].

For example, if the SPL at 30cm is available, the SPL at 1m can be computed as:

$$SPL_{1m} = SPL_{30cm} + 20 \cdot \log_{10}\left(\frac{1}{0.3}\right) \tag{2.8}$$

where the distance d_0 is equal to 30cm and the distance d is equal to 1m. So, from the value of SPL at 30cm it is possible to compute the SPL at 1m by operating a subtraction of a constant value.

2.4.2 Spectral domain parameters

In this section the spectral domain vocal parameters explored in detail are:

- *Harmonic to Noise Ratio (HNR)*

  It is a measure of the harmonicity of the signal and it is expressed in dB.

  It can be computed from the autocorrelation of the signal. In particular, considering a signal in the time domain $x(t)$ and its autocorrelation function $r_x(\tau)$, where τ is the lag, the autocorrelation function is defined as:

$$r_x(\tau) = \int x(t)x(t+\tau)\,dt \tag{2.9}$$

This function can be denominated periodic if, in addition to the global maximum for $\tau = 0$, it has also global maxima outside zero. In this case, the period T_0 exists and all the maxima are placed at a distance equal to integer multiples n of T_0, with $r_x(nT_0) = r_x(0)$.

If no global maxima are present outside zero, there can be local maxima. Considering the case in which the highest local maxima has the height $r_x(\tau_{max})$ at a lag τ_{max}, the signal can be considered to have a periodic part. Consequently, the harmonic strenght R0 is a value between 0 and 1 and equal to the local maximum $r'_x(\tau_{max})$ of the normalized autocorrelation:

$$r'_x(\tau) = \frac{r_x(\tau)}{r_x(0)} \tag{2.10}$$

A signal $x(t)$ like this can be generated taking a periodic signal $H(t)$ characterized by a period T_0 and a noise signal $N(t)$ and summing them. These two signals are uncorrelated, like deduced from equation (2.1).

The autocorrelation of the total signal is equal to the sum of the autocorrelation of the periodic signal and the autocorrelation of the noise signal. For zero lag: $r_x(0) = r_H(0) + r_N(0)$. If the noise is white and so there is no correlation of this signal with itself, there is a local maxima at a lag $\tau_{\max} = T_0$ with a height $r_x(\tau_{\max}) = r_H(T_0) = r_H(0)$. Due to the fact that the autocorrelation at a zero lag represents the power of the signal, the normalized autocorrelation at τ_{max} is the relative power of the harmonic component of the signal. Indeed, its complement is the relative power of the noise component of the signal [12].

$$r'_x(\tau_{\max}) = \frac{r_H(0)}{r_x(0)} \tag{2.11}$$

$$1 - r'_x(\tau_{\max}) = \frac{r_N(0)}{r_x(0)} \tag{2.12}$$

The HNR can be defined as [12]:

$$HNR = 10\log_{10}\left(\frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})}\right) \tag{2.13}$$

- *Absolute Jitter*

  It is a measure of the cycle-to-cycle variation of the fundamental frequency. It is expressed in microseconds [13].

  It can be computed as:

$$Absolute\ Jitter = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_0^{(i)} - T_0^{(i+1)}\right| \tag{2.14}$$

  where $T_0^{(i)}$ with i =1,2,...N is the extracted data of the pitch period and N is the number of extracted pitch periods.

  The absolute jitter, being an absolute measure, is dependent and very sensitive to the fundamental frequency and so, if there are errors in the estimation of the f0, the jitter value can be overestimated or underestimated even significantly. In particular, lower absolute jitter values are associated with higher fundamental and higher absolute jitter values are associated with lower fundamental frequencies.

- *Spectral Slope*

  It is a measurement that describes the change in energy of a signal across frequencies. This is possible by comparing the low frequency spectral energy with the high frequency one. It is often expressed in dB per octave.

  It can be computed as:

$$Spectral\ Slope = 20 \cdot \log\left(\frac{E_1}{E_2}\right) \tag{2.15}$$

  where E_1 is the energy of the signal at the low frequency and E_2 is the energy of the signal at the high frequency.

- *Spectral Tilt*

  It is a concept closely related to spectral slope, but generally refers to the overall slope of the entire frequency spectrum of a signal.

  Indeed, it is often computed as the slope of the trend line that approximates the Long Term Average Spectrum (LTAS).

  The LTAS is proven to be useful in the evaluation of the voice quality and it is computed as the average of the spectrum of the signal. It allows to understand the distribution of the energy in the frequency domain which is useful to distinguish between a healthy and a pathological voice [15].

2.4.3 Cepstral domain parameters

Speech analysis using time domain parameters alone is proved to be insufficient. Infact, temporal parameters such as jitter and shimmer are characterized by the impossibility of carrying out a reliable individuation of the cycle boundaries.
This individuation is trustworthy for signals that are periodic, but not for the less periodic signals that distinguish for example the dysphonic individuals. Futhermore, they are not reliable with very disturbed signals.
The cepstal domain parameters are used to overcome these limitations [16].
The cepstrum is obtained applying the Fourier Transform to the spectrum. In particular, considering a signal s(t), the cepstrum can be computed as:

$$C_r(q) = \mathcal{F}\left\{\log\left(|\mathcal{F}\{s\}|^2\right)\right\} \tag{2.16}$$

where $S^2(f)$ is the power spectrum of s(t) and is expressed as [17]:

$$S^2(f) = \mathcal{F}\{E[s(t) \cdot s^*(t-\tau)]\} \tag{2.17}$$

In this section the cepstral domain vocal parameters explored in detail are:

- *Cepstral Peak Prominence (CPP)*

  It is a measure in dB of the amplitude of the peak in the cepstral domain normalized for the amplitude of the whole signal.

  This normalization is done using a linear regression line that rapresents the relation of the quefrency with the cepstral-magnitude.

  Therefore, the difference (in dB) between the peak in the cepstral domain and the regression line under is computed and it rapresents the CPP [18].

- *Cepstral Peak Prominence Smoothed (CPPS)*

  It is a smoothed version of the CPP. It is computed considering two smoothing steps before the calcutation of the CPPS: a smoothing step over time averaging the cepstra in the time domain and a smoothing step over quefrency averaging the cepstral magnitude in the quefrency domain [18].

  Fig. 2.10 shows the various steps of the computation of the CPPS. In the first part the unsmoothed cepstrum is represented, then the cepstrum smoothed over time and finally the cepstrum smoothed over quefrency.

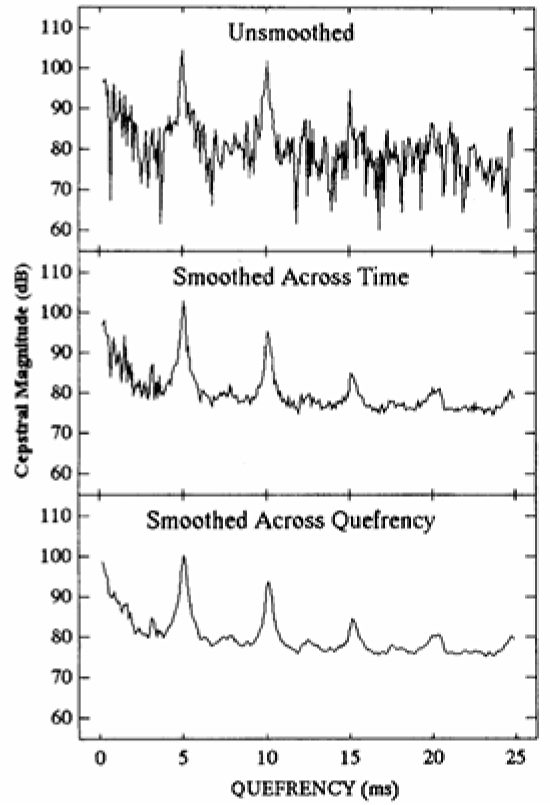

Figure 2.10: Smoothing steps for CPPS computation [18].

2.5 Vocal Indexes

The vocal indexes are used to evaluate the voice quality and to detect the presence of pathologies and the severity of them.
The vocal indexes used in this thesis are the Acoustic Voice Quality Index (AVQI) and the Warning Score (WS), described in the following subsections.

2.5.1 Acoustic Voice Quality Index

The Acoustic Voice Quality Index (AVQI) is a multidimensional index that is used to evaluate the voice quality.
It comes from the need to combine both continous speech and sustained vowel recordings in the extraction of a single index. This is due to the fact that these types of recordings are complementary and they can provide different information about the voice quality.

The sustained vowel recording is characterized by [19]:

- The stability of the phonation.
- The absence of unvoiced segments.
- The absence of prosodic variations both in frequency and intensity.
- The absence of the influence of execution speed, pauses during the execution and phonetic environment.
- The absence of fast changes in the vocal tract configuration.
- The possibility to be produce in a controlled way and with less effort.

The continous speech recording is also informative for several reasons which are useful to conduct a more in depth voice quality evaluation [19]:

- Possibility to detect vocal irregularities.
- It expresses different types of vocal dysphonia.
- The dysphonia can be more evident than in the sustained vowel recording.
- It is rapresentative of the daily use of the voice.

For these reasons, the AVQI is computed considering both the sustained vowel and the continous speech recordings.
In particular, three seconds of the sustained vowel /a/ and voice segments of a phonetically balanced text are used. In this project, three seconds of the middle part of the sustained vowel recording ('sv') and a part of the reading passage ('lect') are selected; these two recordings are concatenated to obtain a single file that is used to compute the AVQI.
The equation of the AVQI is based on the combination of different vocal parameters extracted from the voice recordings: time domain parameters, spectral domain parameters and cepstral domain parameters.
Specifically, the computation of the AVQI includes the Smoothed Cepstral Peak Prominence (CPPS), the Harmonics-To Noise ratio (HNR), the Shimmer Local (Shim), the Shimmer Local in dB (ShdB), the Slope (Slope), the Tilt (Tilt) and it is performed as follow [20]:

$$\begin{aligned} \mathrm{AVQI} = \Big[&4.152 - (0.177 \cdot \mathrm{CPPS}) - (0.006 \cdot \mathrm{HNR}) - (0.037 \cdot \mathrm{Shim}) \\ &+ (0.941 \cdot \mathrm{ShdB}) + (0.01 \cdot \mathrm{Slope}) + (0.093 \cdot \mathrm{Tilt})\Big] \cdot 2.8902 \end{aligned} \tag{2.18}$$

The use of this vocal index is proven to be precise and reliable in the evaluation of the voice quality and in the detection of the presence of pathologies; it is very sensitive to voice modifications which can occur after a vocal therapy; it has a strong relevance at present and finally it is characterized by a high ability to assess phonetic differences [21].
A lower AVQI score corresponds to a healthier person than one characterized by a higher score.
In the case of the subjects treated in this thesis, the AVQI is used to evaluate the gravity of the multiple sclerosis pathology, to detect eventual changes in the voice quality after the LSVT therapy as well as to evaluate the non-inferiority of the telerehabilitation treatment.
The threshold that delimits pathological and non-pathological patients is set at 2.35: subjects below the threshold are healthy and patients above it are affected by multiple sclerosis.
The computation of the AVQI is performed in Praat, VOXplot and Matlab and it is described in the chapter 2.7.
Fig. 2.11 and Fig. 2.12 show respectively an example of the output of the AVQI computation in Praat and in VOXplot.
The calculation of the AVQI is conducted taking into account all the three sustained vowels recordings available for each subject. This allows to have a more reliable and robust evaluation of the voice quality of the patient, considering the variability of the voice in the different recordings. Indeed, the first recording usually has poorer performances because the subject has to get used to the task, the second one is usually optimal and finally vocal tiredness may occur when performing the third repetition.
For these reasons, three different sounds renamed 'avqi1', 'avqi2' and 'avqi3' are obtained for each subject, each one corresponding to the concatenation of the lecture passage and of a different recording of the sustained vowel.
The vocal parameters are computed for each of these three files, they are then averaged and the AVQI is computed on the average values.
This procedure opens the possibility of evaluating the uncertainty of the AVQI score and it is calculated following the approach introduced in [22].
This uncertainty takes into account both the variability of the voice of the subject during the esecution of the sustained vowel and the type of therapy followed. It allows to compare the AVQI score of the two groups of subjects that are treated with the two different therapies and to evaluate the non-inferiority of the telerehabilitation treatment.
Considering the class of subjects in telerehabilitation, the uncertainty can be computed following various steps:

1. Calculation of the mean value x_{i_tele} and of the standard deviation u(x_{i_tele})

for each subject.

2. Calculation of the mean value M_{tele} of all the mean values:

$$M_{tele} = \frac{1}{N}\sum_{i=1}^{N} x_{i\_\text{tele}} \tag{2.19}$$

where N is the number of telerehabilitation subjects, in this case N=10.

3. Calculation of the type A standard uncertainty $u_A(M_{tele})$:

$$u_A(M_{tele}) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{i\_\text{tele}} - M_{tele})^2} \tag{2.20}$$

4. Calculation of the type B standard uncertainty $u_B(M_{tele})$:

$$u_B(M_{tele}) = \frac{1}{N}\sqrt{u^2(x_{tele}) + ... + u^2(x_{N\_\text{tele}})} \tag{2.21}$$

5. Calculation of the combined standard uncertainty $u(M_{tele})$:

$$u(M_{tele}) = \sqrt{u_A^2(M_{tele}) + u_B^2(M_{tele})} \tag{2.22}$$

6. Calculation of the expanded uncertainty $U(M_{tele})$:

$$U(M_{tele}) = k \cdot u(M_{tele}) \tag{2.23}$$

where k is the coverage factor and it is equal to 2.

7. Repetition of the steps form 1 to 6 for the class of subject in presence. $U(M_{pres})$ will be obtained and it can be compared with $U(M_{tele})$.

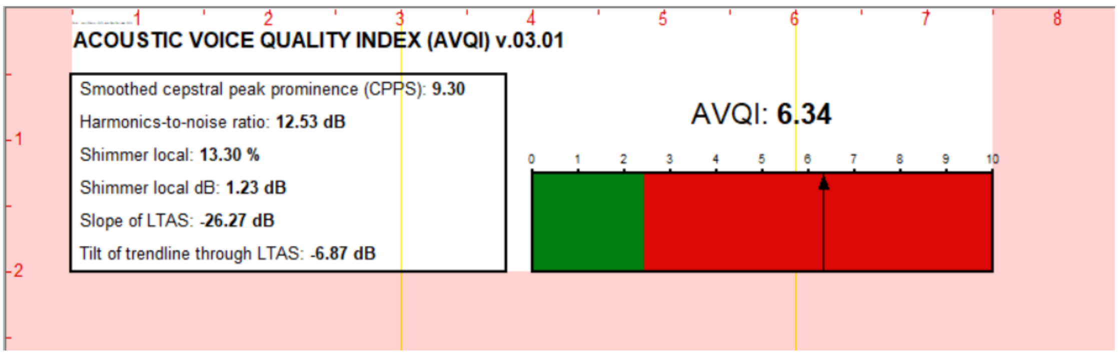

Figure 2.11: Example of the output of the AVQI computation in Praat.

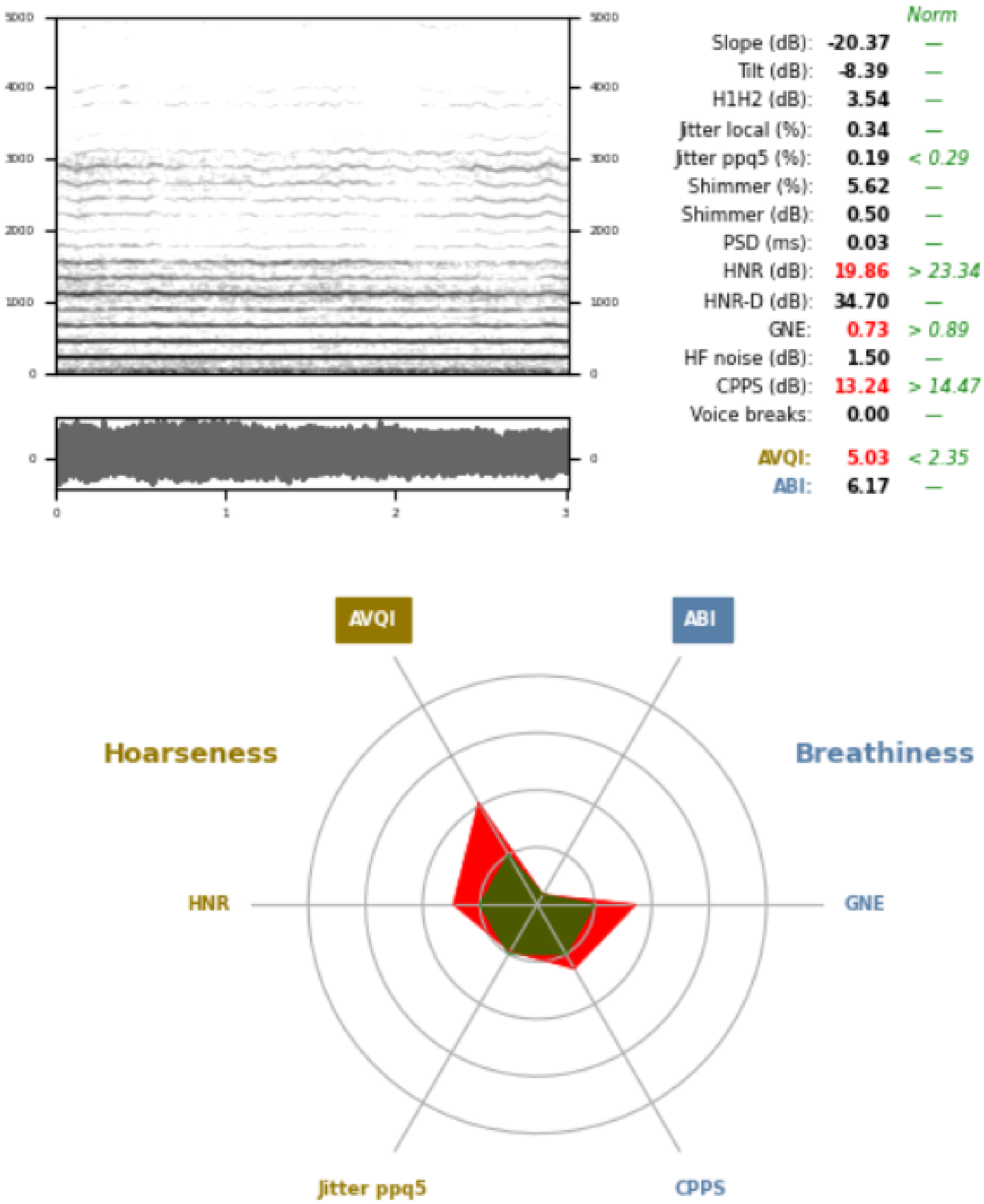

Figure 2.12: Example of the output of the AVQI computation in VOXplot.

The computation of the AVQI is conducted only using the microphone in air. Indeed, it returns the various recordings that can be processed in order to extract the parameters of interest.
The computation is not possible using the Vocal Holter device because it does not provide the Slope, the Tilt and the Shimmer in dB.
In addition to this, the contact microphone only makes the parameters for the sustained vowel available. Indeed, the parameters for the reading passage are not calculated only on the specific recording, but on a file that is the concatenation of the reading and free speech recordings. This makes it impossible to have vocal parameters only for the selected reading sentences.
At the same time, this situation can be transposed for the sustained vowel because it is not possible to have information only for the three selected seconds of the sustained vowel.

2.5.2 Warning score

The warning score (WS) is a vocal index that is used to evaluate the vocal condition of the subject and to detect the presence of pathologies.
It is computed considering a sustained vowel recording for a time interval between 6s and 12s and it involves the extraction of four vocal parameters: the local jitter (%), the local shimmer (%), the CPPSmedian (Cepstral Peak Prominence Smoothed median, dB) and the CPPSstd (Cepstral Peak Prominence Smoothed standard deviation, dB).
Indeed, this index is defined on the basis of the parameters given by the vocal holter device, that are combined in order to understand the presence or absence of a vocal pathology and, if so, its severity.
In particular, different cut-off values are defined so that the value of each extracted parameter corresponds to a score equal to 0 (range of voice not reliable), +1 (patological voice) or -1 (healty voice).
The cut-off values, and so the contibution of the local jitter and of the local shimmer, are defined as follow:

- jitter<0.31% (shimmer<2.37%) for the healthy voice.
- jitter>0.43% (shimmer>2.55%) for the patological voice.
- jitter in the range (0.31÷0.43)% and shimmer in (2.37÷2.55)% for the not reliable voice.

The cut-off values, and so the contibution of the CPPSmedian and of the CPPSstd, are defined as follow:

- CPPSmedian>19.7dB (CPPSstd<0.9dB) for the healthy voice.

- CPPSmedian<18dB (CPPSstd>1.3dB) for the patological voice.
- CPPSmedian in the range (18.0÷19.7)dB and CPPSstd in (0.9÷1.3)dB for the not reliable voice.

The scores obtained for each of the four parameters are then added, obtaining the WS value.
It ranges from -4 to +4 where the interval between -4 and -1 correspond to a healthy voice, the interval between +4 and +1 correspond to a patological voice and a WS equal to 0 indentifies a subject that cannot be classified into one of the two categories [23].
The WS in this study is computed both with a microphone in air and with a contact microphone.
The computation with the contact microphone, as said above, is possible because the Vocal Holter device returns the parameters necessary for its calculation. Since the parameters are already available as output, it is only necessary to apply the cut off values and calculate the corresponding WS score.
Regarding the microphone in air, the parameters necessary for the calculation of the WS are not directly available and they are obtained from the signal of the sustained vowel recording with an appropriate software.
In particular, this is not possible using Praat and VOXplot.
Indeed, Praat did not integrate the possibility to extract the CPPSmedian and the CPPSstd, but only the CPPSmean that is not included in the WS computation.
Infact, the CPPS has recently been integrated into Praat, unlike the CPP parameter which has been present for a longer time and for which it is possible to obtain other values in addition to the mean one, such as the standard deviation, the quantile, the maximum, the minimum and the median absolute deviation.
Regarding VOXplot, it only allows to extract the CPPSmean.
For these reasons, the extraction of the WS with the microphone in air is computed only using Matlab and not Praat and VOXplot.
What should be underlined is that the parameters are extracted on the entire file of the sustained vowel and not on a file of a selected fixed duration. In this way, the values extracted are comparable with those of the contact microphone which are obtained from the entire recording and not from one with a fixed selected duration. Indeed, for the Vocal Holter there are no intelligible signals, so it is not possible to select a part of the recording.
The two microphone configurations allow to evaluate the reliability of the contact microphone compared to the microphone in air and to understand if the contact microphone can be used in the future to evaluate the voice quality of the subjects in terms of WS.

2.6 Pre-Processing

The pre-processing is done simultaneously for all the types of available data: three sustained vowels, a reading passage and a monologue recording.
All the recordings are available for each subject, so there aren't problems of missing data. Therefore, there is no need to make a selection of which data it is possible to use and which not. Furthermore, the data are very clean with no background or overlapping noise, due to the fact that a lot of care was taken in acquiring them. Consequently, there is no need to discard any audio for this reason.
The first part of pre-processing is conducted in Audacity (version 3.2.5), a multi-track and multi-platform audio editing software. It is adopted to cut the parts of silence at the beginning and at the end of the recordings of each of the three sustained vowel.
Then, Audacity is used to select only three seconds, characterized by stability of the fundamental frequency, of the sustained vowel recordings. This is necessary because the extraction of the Acoustic Voice Quality Index requires a fixed duration of the vowel. Three files renamed 'sv1' (sustained vowel), 'sv2' and 'sv3' are obtained.
The software is then used to separate the reading passage and the monologue recording which are provided in a single file, removing also the silences before and after them. Also in this case a selection of parts of these recordings is necessary.
For the monologue, a part of about one minute is selected, starting from the beginning of the recording. This decision is done because some recordings are very long and one minute is considered a good compromise between the need to have a sufficient length of the files to evaluate vocal parameters and the necessity to have a manageable data. The file obtained is renamed 'mon' (monologue).
Regarding the lecture, the same sentences of the passage ' Notturno' are selected for all subjects. This arises from the necessity to have a fixed passage for the future computation of the Acoustic Voice Quality Index, that is also shorter of the complete passage. The file obtained is renamed 'lect' (lecture).
All the recordings are appropriately divided into folders distinguishing patient and acquisition time (T0 or T1).
The output files of Audacity in .wav format are then imported in Praat, VOXplot and Matlab in order to be appropriately pre-processed.

2.6.1 Matlab Pre-Processing

All the recordings are loaded in Matlab (R2022b) to be pre-processed, in order to distinguish between voiced and unvoiced frames.
The pre-processing is conducted in the same way for the sustained vowel and the monologue and lecture recording, but the difference can be found in the fact that the sustained vowel recording is divided into pseudoperiods using an autocorrelation

algorithm, while the monologue and lecture recordings are divided into frames of 2048 samples which correspond to a time interval of 46 ms.
The first step is to initialize various parameters such as number of channels, threshold for distinguish voiced and unvoiced frames, minimum number of frames to discard before accepting a new valid frame, the maximum and minimum value of the fundamental frequency on the basis of the gender of the subject.
At this point, the additive noise present in the signal is verified, in order to understand if it is low enough; then vertical resampling is performed.
Then, the mean of the signal is removed and on the signal an amplitude normalization is performed.
After all this steps, the removal of the silent frames is conducted. For the free speech recordings the removal of silence is essential. Infact, even if the recordings are already cleaned from the silences at the beginning and at the end, there are still some inside the recordings between the sentences due to pauses made by the subject. For the sustained vowel this is not necessary because the vowel is a continous sound and there are no silences within it. Indeed, a specific request of the task is to emit the vowel continuously.
With the aim to remove the silences, an RMS-based (Root Mean Square) technique is used to distinguish between voiced and unvoiced frames. In particular, the RMS is computed for the specific frame and a threshold is set to the half of the RMS of the whole signal: if the RMS of the frame is higher than the threshold, then it is considered as a non-silence frame (voiced). Otherwise, if the value of the frame is under the threshold then it is considered a silence frame (unvoiced) and it is rejected.
At this point, in order to assess the harmonic quality of the frames, the HNR (Harmonic-to-Noise Ratio) value is computed for each frame labelled before as voiced and a threshold is set at 0dB to distinguish between harmonic and non-harmonic frames. In particular, the frames that have a value of HNR higher than the 0 dB threshold are selected and considered as harmonic.
Finally, it is possible to individuate an appropriate fundamental frequency checking the frequency jump.
The output of this pre-processing is a set of files characterized only by frames with voiced parts and that are at this point ready to be processed in order to extract the vocal parameters and the vocal indexes.

2.6.2 Praat and VOXplot Pre-Processing

The pre-processing carried out in Praat and VOXplot is completely comparable. Indeed, VOXplot is based on reliable algorithms developed in Praat environment [11] and there is no possibility of making changes to its codes but only of extracting the preset parameters. VOXPlot is used to extract the vocal parameters and

the vocal indexes from the recordings of the sustained vowels and the monologue recordings following the rules reported in the VOXplot site [11].
Before the pre-processing, unlike Matlab, there is no division into frames for the free speech recordings or into pseudoperiods for the sustained vowels during the first phase of the analysis. The division in frames, for example for the removal of the silences, is included in the appropriate Praat functions that will be used.
Firstly, a sound named 'onlyVoice' with a duration of 0.001 s and with the same sampling frequency of the original sound is created.
Then, the silences are identified and removed from the signals using the 'ToTextGrid' function. In this object have to be defined the silence threshold (dB), the minimum duration of the silence interval (s) and the minimum duration of the sounding interval (s).
The minimum duration of the silence interval is the minimum time that a silence interval has to last to be considered as a silence and it is set at 0.046 s. The minimum duration of the sounding interval is the minimum time that a sounding interval has to last to be considered as a sounding interval and it is set at 0.046 s. The default value in both cases is 0.1 s but it is reduced to 0.046 s in order to consider the same frames of the signal as in Matlab.
The silence threshold defines the maximum value of the intensity with respectively to the maximum intensity value of the signal. In particular, this value is set at -25 dB (default value) and this means that the threshold is calculated as the difference bertween the maximum intensity (imax) value of the signal and the silence threshold, -25 dB in this case. The obtained value is the one used to check if a frame is a silence or a sounding frame: if the intensity value of the frame is lower than the threshold, then it is considered a silence frame.
Two labels (sounding and silent) are used to label the sounding and the silence intervals. The 'sounding' intervals correspond to one and the 'silent' intervals correspond to zero.
After this labeling, some of these parts of the signal labelled as 'sounding' or 'silent' are removed. In particular, the sounding intervals with a duration smaller than minimum sounding interval duration and the silent intervals with a duration smaller than the minimum silent interval duration are excluded.
Finally, all the frames labelled as 'sounding' are selected and concatenated in a new sound file renamed 'onlyLoud'.
It is possible using the 'Extract intervals where' function, a Praat object that allows to extract a part of the signal from the original sound.
In particular, it is necessary to specify that the intervals to be extracted are the 'sounding' intervals, so the intervals marked with a one. Intervals marked as 'silence' are therefore completely discarded.
At this point, the global power of the sound 'onlyLoud' is computed and a voiceless threshold is set at 30% of this value. This threshold is used to select the voiced

frames from the 'onlyLound' sound.
The signal 'onlyLound' is analyzed in frame of 0.03 s: the power of this frame (partial power) is computed and it is compared with the voiceless threshold. If the power of the frame is higher than the threshold, then the frame is considered voiced.
If this condition occurs, the segment is further analyzed for the zero crossing rate. If the zero crossing rate is undefined or is less than 3000 number of crossings per second, the segment is considered vocal and added to the 'onlyVoice' sound created previously.
The output of this pre-processing is a set of files characterized only by frames with a vocal content and that are ready to be processed in order to extract the vocal parameters and the vocal indices.

2.7 Parameters and Indexes Extraction

In this chapter, when dealing with parametric extraction, the focus is on the processing of recordings operated in Praat and Matlab.
Infact, VOXplot does not integrate the possibility of modifying its codes and can only be used to extract the set parameters. VOXplot is instead employed to check the compatibility of its results with those provided by Praat, as will be seen in the next chapter.
The extraction of the parameters useful for calculating the vocal indexes is firstly analyzed. The vocal indexes involved in the study are the Acoustic Voice Quality Index (AVQI) and the Warning Score (WS).
The parameters involved in the calculation of the AVQI are the Smoothed Cepstral Peak Prominence (CPPS), the Harmonics-To-Noise ratio (HNR), the Shimmer Local, the Shimmer Local in dB, the Spectral Slope and the Spectral Tilt.
In a first version of the parametric processing (*'Version 1'*) of the AVQI, the code presented in the appendix of [21] is used to obtain the values from Praat.
This processing version produces satisfactory results but at the same time various changes (*'Version 2'*) are made in Praat, in order to express with greater precision and clarity the parametric implementation of HNR and CPPS.
For what concerns Matlab, a script specifically developed to analyze the concatenation of a sustained vowel and a reading passage recordings is employed, following the same rules of the parameters extraction conducted in Praat.
Below, the parametric implementation of *'Version 2'* in Praat is explored in depth, highlighting any variation with the *'Version 1'* of the processing, in order to understand the way of operating of this sofwtare application:

- *Spectral Slope*

  The Spectal Slope is implemented as the difference between the energy in the low frequency range (from 0 to 1000 Hz) and the energy in the high frequency range (from 1000 to 10 000 Hz) of the long-term average spectrum (LTAS).

  In order to compute the Spectral Slope, the object *'To Ltas'* is used to calculate the LTAS of the signal. Then the Spectral Slope is calculated using the object *'Get slope'* to which various parameters are passed: the minimum frequency (0 Hz) and the maximum frequency (1000 Hz) of the low band, the minimum frequency (1000 Hz) and the maximum frequency (10 000 Hz) of the high band, the type of averaging method ('energy'). This implementation in Praat has no differences with the *'Version 1'* one.

- *Spectral Tilt*

  The Spectral Tilt is implemented as the difference between the the energi in the the low frequency range (from 0 to 1000 Hz) and the energy in the high frequency range (from 1000 to 10 000 Hz) of the trendline through the long-term average spectrum.

  In order to compute the Spectral Tilt, the object *'Compute trend line'* is used: it allows to obtain the trend line of the LTAS, specifying the minimum frequency (0 Hz) and the maximum frequency (10 000 Hz) of the band to be considered. Then the object *'Get slope'* is used to calculate the Spectral Tilt, passing the same parameters as for the Spectral Slope. This implementation in Praat has no differences with the *'Version 1'* one.

- *Cepstral Peak Prominence Smoothed (CPPS)*

  The CPPS is implemented using the object *'To PowerCepstrogram'* applyed to the sound in order to calculate the power cepstrogram of the signal, specifying the pitch floor (60Hz), the time step that is the distance between two frames (0.002s), the maximum frequency (11025Hz) useful because the signal is resampled to twice this value, the value of the pre-emphasis filter (1Hz).

  Then the object *'Get CPPS'* is applyed to the obtained power cepstrogram. It is used to compute the CPPS, specifying various input: no subtraction of the trend before smoothing, the time averaging window (0.014s), the quefrency averaging window (0.001s), the peak search pitch range (60-330Hz), the tolerance (0.05), the interpolation so how the amplitude and position of the peak are evaluated (Parabolic), the quefrency range for which the amplitudes will be modelled by a straigth line (0.001-end of the quefrency range), the trend type that defines how to model the cepstrum (Straight), the fit method that defines how the line that models the cepstrum background is computed (Robust).

In this version (*'Version 2'*) there are changes compared to the *'Version 1'* of the processing in both objects *'To PowerCepstrogram'* and *'Get CPPS'*.

In particular, in *'Version 1'* in *'To PowerCepstrogram'* object, the maximum frequency is set at 5000 Hz (default value), consequently the signal is resampled at 10 000 Hz. This value of resampling is not commonly used in the literature, for this reason it is set at 11025 Hz guaranteeing a resample of the signal at 22050 Hz. In addition to this, the default value of the pre-emphasis filter is 50 Hz, but it is set at 1 Hz in order to perform a null pre-emphasis. Infact, its action does not produce significant variation of the results, so it is not employed.

Considering the *'Get CPPS'* object, the changes are related to quefrency range for which the amplitudes will be modelled by a straigth line. The values of the range in *'Version 1'* do not correspond to the full range of the quefrency and they are equal for every signal. In *'Version 2'* the full range of the quefrency is selected using the object *'Get end quefrency'*, that allows to obtain the end of the quefrency range, different for each signal. The selection of the full range is done in order to better estimate the CPPS.

- *Local Shimmer*

  The Local Shimmer is computed as the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude.

  It is implemented using the object *'To PointProcess(periodic,cc)'* applyed to the sound and specyfing the pitch floor (50 Hz) and the pitch ceiling (400 Hz), in order to evaluate the periodicity of the signal. Infact, a point process is obtained and it is a sequence of points t_i defined on all the time domain [t_{min},t_{max}]. This object seeks the near locations of high amplitude using the cross-correlation method and it returns a sequence of points.

  The latter is superimposed to the signal and then the object *'Get shimmer (local)'* is used to calculate the Local Shimmer, specyfing various parameters: the time range on which it must be calculated (0,0 that means on all the signal), the lenght of the shortest period (0.0001s), the length of the longest period (0.02s), the maximum period factor (1.3) and the maximum amplitude factor (1.6). In particular, the minimum and maximum lenght of the period refer to the time window for searching consecutive periods. This implementation in Praat has no differences with the *'Version 1'* one.

- *Local Shimmer in dB*

  The Local Shimmer in dB is computed as the average absolute base-10 logarithm of the difference between the amplitudes of consecutive periods, multiplied by 20.

It is implemented following the same steps as the Local Shimmer, but instead of using the object *'Get shimmer'*, the specific object *'Get shimmer(local dB)'* is used. The parameters that are given as input to this object are the same of the Local Shimmer. This implementation in Praat has no differences with the *'Version 1'* one.

- *Harmonic-to-Noise Ratio (HNR)*

  The HNR is implemented in Praat using the object *'To Harmonicity (ac)* applyed to the sound in order to evaluate the periodicity of the signal using an autocorrelation method. Various parameters are specified in this object: the time step (0.01 s), the pitch floor (75 Hz) that defines the lenght of the analysis window, the silence threshold (0.1) which defines as silent the frames whith amplitudes under this value, number of periods per window (4.5).

  Then the object *'Get mean'* is used to calculate the HNR mean, specyfing the time range on which it must be calculated (0,0 that means on all the signal).

  In this version (*'Version 2'*) there are changes compared to the *'Version 1'* of the processing in Praat.

  In particular, in (*'Version 1'*) the object *'Voice report'* is used to calculate various vocal parameters and then the HNR is extracted using the *'extractNumber'* function, specifying the desidered parameter. Instead, in (*'Version 2'*) a specific object for HNR extraction that uses the autocorrelation method is employed. This is done because it is not possible to understand the operations implemented in the Voice Report object; therefore, it was preferred to use an object specifically implemented for HNR extraction and that correctly employs the autocorrelation method.

All the parameters are extracted from a file made up of the concatenation of three seconds of a sustained vowel and a piece of reading of a phonetically balanced recording.
Furthermore, having three vocal recordings available, the parameters are extracted from three different files corresponding to the concatenation of the lecture passage and of a different recording of the sustained vowel. The vocal parameters are computed for each of these three files, they are then averaged and the AVQI is computed on the average values.
The AVQI is computed in Praat and Matlab using a linear formula that combines the six parameters and presented in the chapter 2.5.

The other vocal index involved in the study is the Warning Score (WS). The parameters involved in the calculation of the WS are the Local Jitter (%), the Local Shimmer (%), the CPPSmedian (Cepstral Peak Prominence Smoothed median,

dB) and the CPPSstd (Cepstral Peak Prominence Smoothed standard deviation, dB).
The WS is only computed in Matlab because some parameters necessary for its calculation are not directly available in Praat, in particular the CPPSmedian and the CPPSstd. Infact, Praat only provides the CPPSmean.
The WS is extracted from the sustained vowel recording and for both the microphone in air and the contact microphone.
After the extraction of the parameters of interest, this index is obtained by assigning a score (0, -1 or +1) to the value of each of the four parameters, based on what is defined in the chapter 2.5. A score that ranges from -4 to +4 is obtained.

In addition to these parameters useful for extracting the vocal indexes, another parameter of interest for the vocal analysis is the Sound Pressure Level (SPL). It is computed because it allows to have an immediate idea of the intensity of the voice, which is of particular interest in patients with multiple sclerosis.
This parameter is extracted from the monologue recording selecting a part of about one minute starting from the beginning of the recording.
It is calculated using the formulas presented in the chapter 2.4. As can be seen from them, the Root Mean Square (RMS) is essential and it can be easily computed in Matlab and Praat.
In particular, Praat uses the object *'Get root-mean-square'* to calculate the RMS of the signal, specifying the time range on which it must be calculated (0,0 that means on all the signal).
A difference between Praat and Matlab in the SPL extraction is that in Matlab a vector with RMS values in different frames is obtained, whereas in Praat a single RMS value is obtained which corresponds to the average value.
Consequently, in Praat only the SPLmean can be evaluated. Instead, Matlab allows to carry out a statistical analysis of this parameter, computing for example mean, median, 5° percentile, 95° percentile and standard deviation of the SPL.
The SPL is not computed in VOXplot because it does not provide the RMS value.

The extraction of the vocal parameters and indexes is conducted for all the subjects and for both the acquisition times (T0 and T1).

2.8 Perceptual Rating Scales: VHI

The perceptual rating scales are used to evaluate subjectively (carried out by the patient) or objectively (carried out by the phoniatrician) the condition of the patients after a treatment.
In this study, a subjective perceptual rating scale, that is the Voice Handicap Index (VHI), is used to understand the perception of the subjects of their voice before the therapy (at T0) and after the therapy (at T1), evaluating their vocal handicap.

The VHI is a questionnaire that is made of 30 questions, divided into three subgroups of questions:

- Functional: it considers the impact of vocal problems on daily life.
- Psychological: it considers the psychological impact.
- Emotional: it considers the perception of the characteristics of vocal emission.

Each answer is evaluated with a score that ranges from 0 to 4, where 0 corresponds to 'never', 1 to 'almost never', 2 to 'sometimes', 3 to 'almost always' and 4 to 'always'. Therefore, each subgroups of questions can have a score from 0 to 40. These are then summed to obtain the total score that ranges from 0 to 120.
The total score is evaluated in the following way:

- 0: normal.
- 1-40: mild alteration.
- 41-80: moderate alteration.
- 81-120: severe alteration.

In this way, the three different aspects of the vocal handicap can be evaluated separately or globally [24].

Chapter 3

Results and Discussions

The results, otained from the analysis conducted in this thesis, are presented and discussed in detail in this chapter.
The results related to the AVQI are presented with the aim to compare the software applications introduced in this work: Praat, Matlab and VOXplot. The purpose is to underline the similarity between Praat and VOXplot and to assess the reliability of Matlab results, making a comparison with those obtained from Praat.
The comparison between the software applications is also conducted in terms of SPL, extracted from Matlab and Praat.
The results related to the WS, extracted both from the in-air microphone and the contact microphone, are also presented. This is done in order to assess the reliability of the contact microphone compared to the in-air one.
Finally, the evaluation of the effectiveness of the LSVT theraphy in terms of vocal improvement for both therapies is conducted, underlining above all the non-inferiority of the telerehabilitation therapy compared to the in-clinic one. This analysis is performed in terms of vocal indexes (AVQI and WS) and in terms of vocal parameters, extracted from both the monologue and the sustained vowel recordings.

3.1 Software analysis to compute AVQI

The Acoustic Voice Quality Index (AVQI), introduced in 2.5.1, is a vocal index extracted from a recording made of the concatenation of three seconds of a sustained vowel /a/ and a reading passage.
In this work, the calculation of the AVQI is conducted considering all the three sustained vowels recordings available for each subject. Therefore, the values of the parameters shown in the following tables are the average values computed

considering three AVQI sounds. The AVQI is obtained using a linear formula that includes the values of the Smoothed Cepstral Peak Prominence (CPPS), the Harmonics-To Noise ratio (HNR), the Shimmer Local (Shim), the Shimmer Local in dB (ShdB), the Slope (Slope), the Tilt (Tilt). These parameters are extracted from Praat, Matlab and VOXplot.

3.1.1 Comparison between Praat and VOXplot

The first comparison, in terms of AVQI, is conducted between Praat and VOXplot. The parameters, and so this vocal index, are extracted from Praat using the code presented in the appendix of [21], corresponding to the *'Version 1* of the elaboration conducted in this thesis.
Also VOXplot uses this code, so the two software are expected to provide comparable results. Infact, VOXplot does not integrate the possibility of modifying its integrated codes and can only be used to extract the set parameters.
The AVQI is computed for each subject at T0 and T1 and the results are expressed in terms of AVQImean $\pm$ one standard deviation, where the average is made between all the patients without a distinction, in this first phase of the analysis, between the in-clinic and the telerehabilitation group.
Table 3.1 and Table 3.2 show the AVQI values and the parameters involved in its calculation extracted from Praat at T0 and T1 respectively. At T0 Praat results in an average value of 6.0$\pm$0.29; at T1 the average value is 5.2$\pm$0.29.
Table 3.3 and Table 3.4 show the AVQI values and the parameters involved in its calculation extracted from VOXplot at T0 and T1 respectively. At T0 VOXplot results in an average value of 6.0$\pm$0.30; at T1 the average value is 5.1$\pm$0.29.
It can be noticed that the two software produce identical results as expected, both at time T0 and T1. This can be seen in Fig. 3.1 and in Fig. 3.2, where the comparison between the results in Praat and VOXplot is shown, respectively at time T0 and T1. Instead, in Fig. 3.3 and in Fig. 3.4 the comparison between the mean AVQI values and the uncertainty bars related to the software application are rapresented, for T0 and T1 respectively.
These results therefore confirm the reliability of VOXplot and allow the use of a practical and easy-to-handle software application, which provides various parameters and vocal indexes in a few moments and effortlessly.
One aspect that needs to be paid attention to, and which can alter the results provided by VOXplot, is the number of channels of the files given as input to this software application. Infact, the vocal signals are acquired with two channels, but the extraction of the parameters requires signals with only one: Praat in its scripts automatically extracts one, while VOXplot does not.
This is why the recordings must be provided in 'mono' format (one channel) and not in 'stereo' format (two channels) as input to VOXplot.

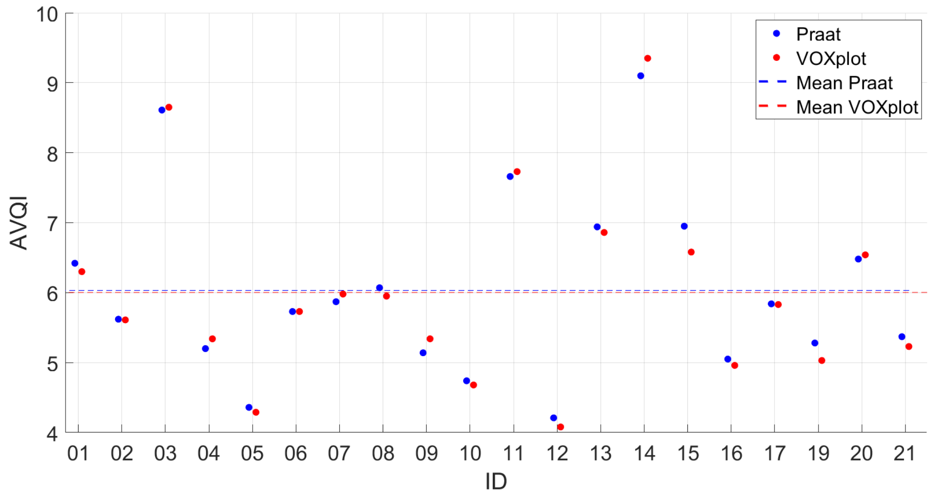

Figure 3.1: AVQI comparison between Praat and VOXplot at T0.

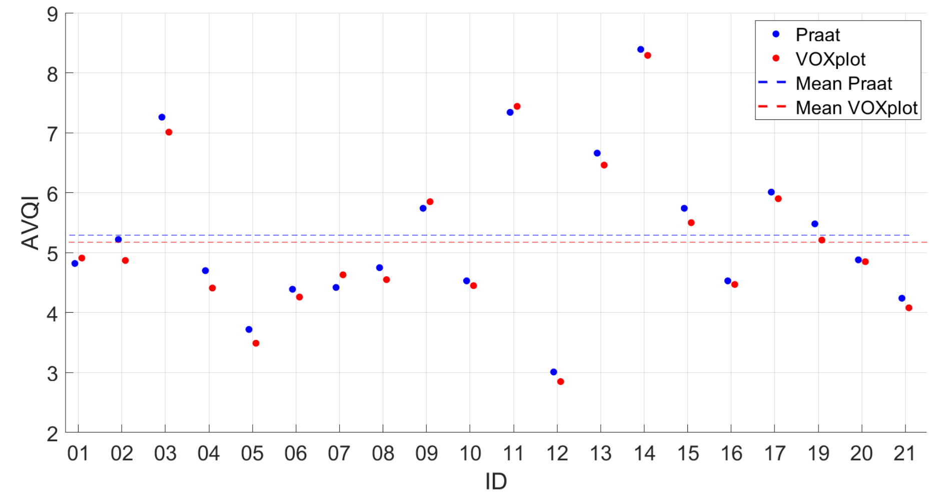

Figure 3.2: AVQI comparison between Praat and VOXplot at T1.

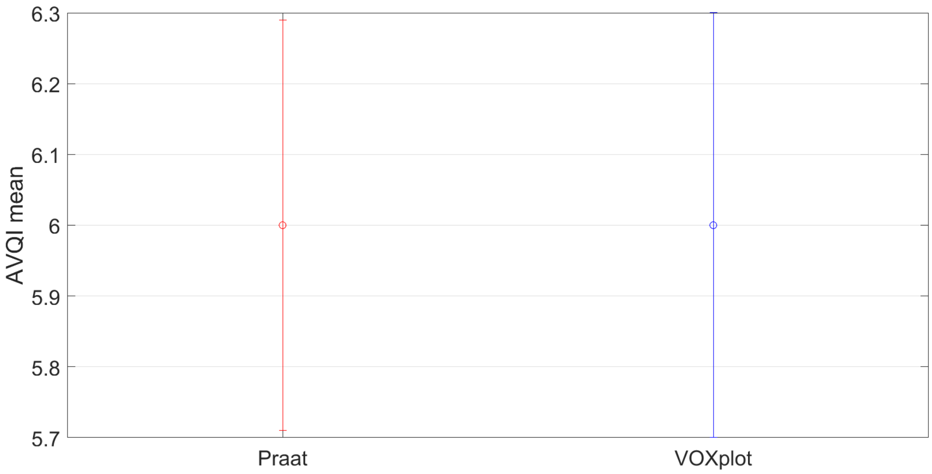

Figure 3.3: AVQI mean comparison between Praat and VOXplot at T0.

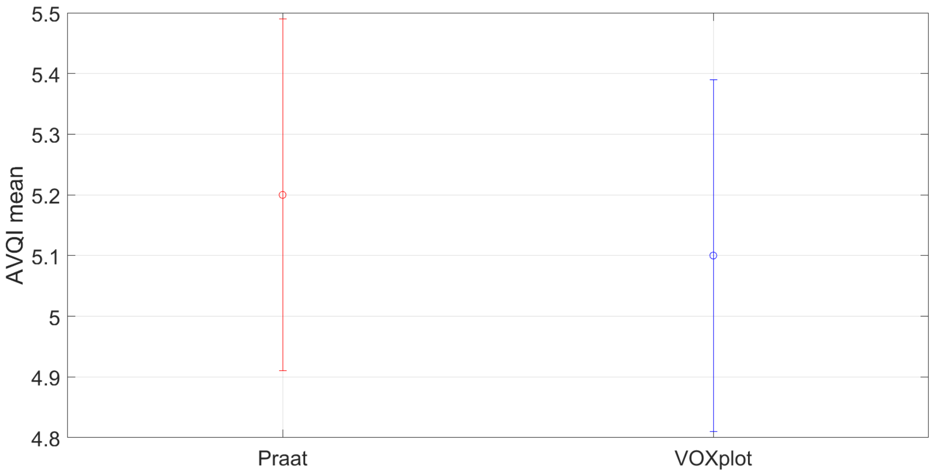

Figure 3.4: AVQI mean comparison between Praat and VOXplot at T1.

	ID	CPPS(dB)	HNR(dB)	Shimmer(%)	Shimmer(dB)	Slope(dB)	Tilt (dB)	AVQI
In-clinic	01	9.2	11.8	13.6	1.2	-24.8	-7.1	6.4
	03	6.5	12.4	13.2	1.2	-29.2	-3.6	8.6
	05	12.1	13.7	10.1	0.9	-20.1	-8.1	4.3
	06	9.7	10.1	12.6	1.2	-18.4	-9.3	5.7
	08	9.2	10.6	14.8	1.3	-23.3	-9.0	6.1
	11	6.0	10.1	15.9	1.4	-28.8	-8.7	7.6
	13	8.5	11.7	14.2	1.3	-26.2	-6.6	6.9
	15	9.1	12.1	10.9	1.1	-23.4	-4.8	6.9
	19	11.8	10.6	9.8	1.0	-11.7	-7.3	5.2
	20	8.5	11.9	12.4	1.1	-27.7	-7.5	6.5
Telerehabilitation	02	10.7	13.4	10.2	1.1	-20.5	-6.7	5.6
	04	10.1	13.5	11.0	1.0	-26.4	-8.4	5.2
	07	9.4	11.0	13.1	1.2	-25.7	-8.2	5.8
	09	10.6	12.1	10.3	0.9	-19.9	-8.1	5.1
	10	11.3	12.8	11.6	1.0	-23.7	-8.3	4.7
	12	12.2	12.5	11.3	1.1	-19.1	-9.4	4.2
	14	4.3	7.9	18.6	1.6	-29.5	-7.6	9.1
	16	10.9	12.5	10.9	1.0	-17.4	-8.2	5.0
	17	9.8	10.5	10.2	0.9	-23.5	-6.0	5.8
	21	10.4	13.2	9.5	0.9	-21.8	-7.2	5.4

Table 3.1: Praat results for the AVQI extraction at T0: *'Version 1'* of the elaboration.

	ID	CPPS(dB)	HNR(dB)	Shimmer(%)	Shimmer(dB)	Slope(dB)	Tilt (dB)	AVQI
In-clinic	01	10.8	13.6	11.1	1.0	-25.5	-8.5	4.8
	03	8.3	12.8	10.1	1.0	-25.3	-4.6	7.2
	05	13.2	14.8	6.6	0.7	-17.6	-7.3	3.7
	06	11.7	11.9	10.6	1.1	-16.5	-10.2	4.3
	08	11.5	14.1	10.7	1.1	-15.0	-9.1	4.7
	09	9.6	10.8	11.3	1.0	-19.6	-8.2	5.7
	10	12.7	12.9	7.2	0.9	-12.5	-7.7	4.5
	11	6.5	11.7	12.8	1.2	-27.3	-8.2	7.3
	12	14.6	16.3	6.5	0.7	-12.0	-8.4	3.0
	13	8.7	12.8	12.8	1.2	-26.1	-7.0	6.6
Telerehabilitation	02	12.1	15.7	9.6	0.9	-11.4	-6.1	5.2
	04	10.6	15.6	8.3	0.9	-26.8	-8.8	4.7
	07	11.5	12.6	10.7	0.9	-24.8	-8.4	4.4
	14	6.3	9.4	16.2	1.5	-25.7	-6.8	8.3
	15	11.2	13.1	8.2	0.9	-17.4	-4.9	5.7
	16	11.4	13.7	8.5	0.8	-15.4	-8.9	4.5
	17	10.1	9.2	13.6	1.2	-22.2	-6.5	6.0
	19	10.9	10.9	12.3	1.1	-17.6	-7.3	5.4
	20	11.1	13.0	10.1	1.0	-25.1	-7.9	4.8
	21	11.9	14.3	8.6	0.8	-10.4	-8.2	4.2

Table 3.2: Praat results for the AVQI extraction at T1: *'Version 1'* of the elaboration.

	ID	CPPS(dB)	HNR(dB)	Shimmer(%)	Shimmer(dB)	Slope(dB)	Tilt (dB)	AVQI
In-clinic	01	9.4	11.4	14.4	1.3	-23.9	-7.4	6.3
	03	6.8	11.6	14.6	1.3	-29.3	-3.8	8.6
	05	12.2	13.8	10.1	0.9	-20.1	-8.0	4.2
	06	9.7	10.5	12.8	1.2	-18.4	-9.4	5.7
	07	9.5	10.1	13.9	1.2	-25.8	-8.2	5.9
	08	9.4	10.6	14.9	1.3	-23.5	-9.1	5.9
	10	11.4	12.6	11.2	1.0	-23.7	-8.3	4.6
	11	6.1	9.8	16.1	1.4	-29.0	-8.8	7.7
	12	12.3	12.8	11.2	1.1	-19.1	-9.5	4.1
	13	8.7	11.6	13.9	1.3	-26.4	-6.7	6.8
Telerehabilitation	02	10.9	13.3	11.9	1.1	-20.6	-6.8	5.6
	04	10.4	11.8	13.6	1.2	-27.2	-8.5	5.3
	14	3.9	7.0	18.5	1.6	-29.6	-7.6	9.3
	15	9.4	12.6	10.5	1.0	-23.6	-5.0	6.5
	16	11.1	13.4	11.1	1.0	-17.5	-8.3	4.0
	17	9.9	10.2	10.5	0.9	-23.6	-6.9	5.8
	19	11.9	11.3	9.6	1.0	-11.8	-7.3	5.0
	20	8.4	12.1	12.8	1.2	-27.7	-7.6	6.5
	21	10.5	14.2	9.3	0.9	-21.8	-7.2	5.2

Table 3.3: VOXplot results for the AVQI extraction at T0.

	ID	CPPS(dB)	HNR(dB)	Shimmer(%)	Shimmer(dB)	Slope(dB)	Tilt (dB)	AVQI
In-clinic	01	10.9	13.6	11.5	1.1	-25.5	-8.5	4.9
	03	8.5	12.9	10.6	1.0	-25.4	-4.7	7.0
	06	11.8	12.2	10.6	1.0	-16.5	-10.3	4.2
	07	11.5	11.8	11.6	1.0	-24.8	-8.4	4.6
	09	9.6	10.8	10.5	0.9	-18.9	-8.4	5.8
	11	6.5	12.2	12.8	1.2	-27.3	-8.2	7.4
	12	14.6	18.0	5.9	0.7	-11.0	-8.5	2.8
	13	8.9	13.1	11.6	1.1	-26.3	-7.0	6.4
	15	11.3	14.4	8.1	0.8	-17.8	-5.8	5.5
	17	10.2	9.0	12.9	1.1	-22.2	-6.6	5.9
Telerehabilitation	02	12.2	16.6	7.4	0.8	-11.4	-6.1	4.8
	04	10.8	15.7	8.0	0.8	-26.8	-8.9	4.4
	05	13.2	15.3	6.2	0.6	-17.6	-7.3	3.4
	08	11.7	14.6	9.2	0.9	-15.0	-9.1	4.5
	10	12.8	13.5	6.4	0.8	-12.2	-7.6	4.4
	14	6.4	10.6	15.6	1.4	-25.5	-6.8	8.2
	16	11.5	14.8	9.2	0.9	-15.4	-8.9	4.4
	19	11.0	11.1	11.7	1.0	-17.7	-7.4	5.2
	20	11.1	13.1	10.8	1.0	-25.2	-8.0	4.8
	21	12.0	15.3	8.7	0.8	-20.4	-8.2	4.0

Table 3.4: VOXplot results for the AVQI extraction at T1.

3.1.2 Comparison between Praat and Matlab

The comparison between Praat and Matlab is conducted in this subsection.
Praat is used as a reference because has been verified in 3.1.1 the compatibility of the results between Praat and VOXplot, therefore it was decided to use only one of the two software applications for the comparison with Matlab.
The parameters and the AVQI are extracted from Matlab using a script specially developed to analyze the concatenation of a sustained vowel and a reading passage recording.
The results in Matlab are shown in Table 3.5 and Table 3.6, for T0 and T1 respectively, and they are expressed in terms of AVQImean ± one standard deviation, where the average is made between all the patients without a distinction, in this first phase of the analysis, between the in-clinic and the telerehabilitation group.
At T0 Matlab results in an AVQI average value of 5.5±0.26; at T1 the AVQI average value is 4.9±0.23.
These Matlab values are compared to those obtained from the second version of the elaboration (*'Version 2'*) of Praat, whose changes compared to the first version (*'Version 1'*) of the elaboration are described in the section 2.7.
The results in Praat are shown in Table 3.7 and Table 3.8, for T0 and T1 respectively.
At T0 Praat *'Version 2'* results in an AVQI average value of 5.6±0.26; at T1 the AVQI average value is 4.9±0.27.
First of all, it is noticeable that the results of this version of the processing in Praat (*'Version 2'*) and the previous one (*'Version 1'*), shown in the previous section 3.1.1, differ very little: the changes were made for greater precision and clarity in the parametric implementation.
Then, it can be noticed that Matlab and Praat provide comparable results in terms of AVQI mean, both at time T0 and T1, as can be seen in Fig. 3.5 and Fig. 3.6.
This is also evident in Fig. 3.7 and in Fig. 3.8, where the comparison between the mean AVQI values and the uncertainty bars related to the software application are rapresented, for T0 and T1 respectively.
It is therefore evident how Matlab and the developed codes are also effective in the evaluation of this vocal index.
This opens up the possibility in the future of integrating these developed codes into the Vocal Holter device, in order to make it a diagnostic tool also in terms of AVQI.
The AVQI computation would be an additional function to those currently available, which only involve the extraction of vocal parameters of interest. This can open up the possibility of immediately and more easily interpreting the patient's clinical situation through the use of a vocal index.

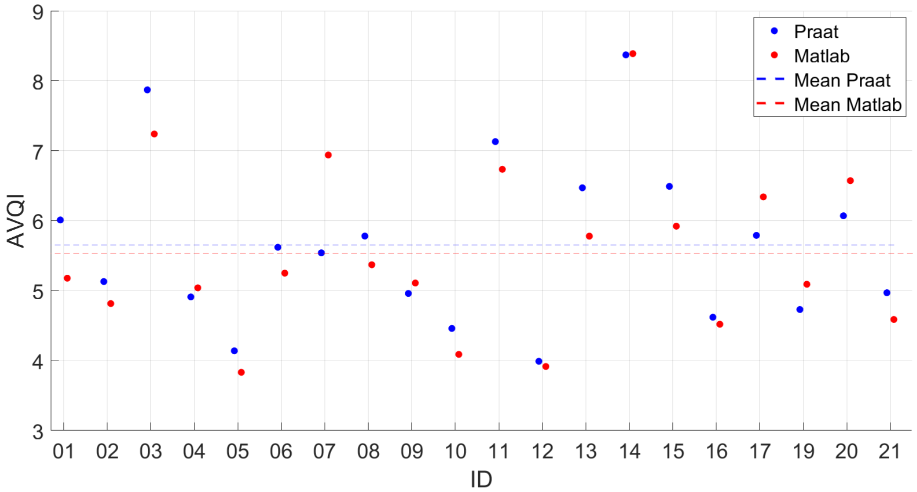

Figure 3.5: AVQI comparison between Praat and Matlab at T0.

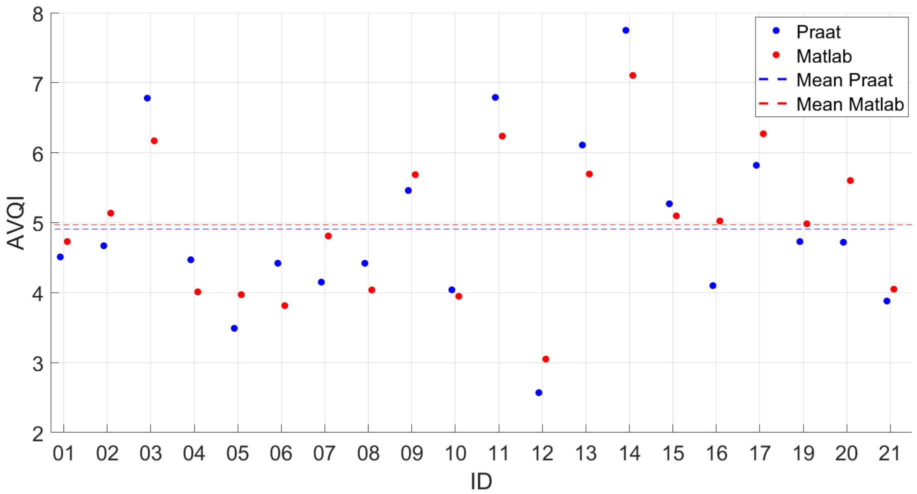

Figure 3.6: AVQI comparison between Praat and Matlab at T1.

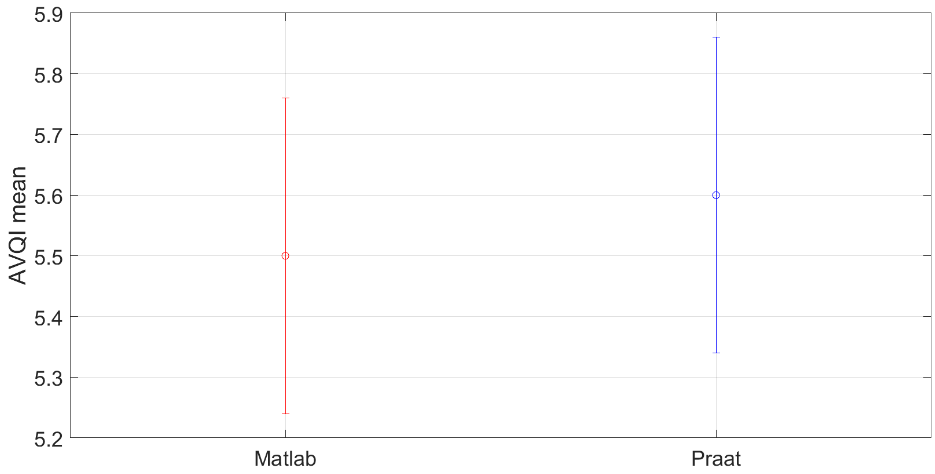

Figure 3.7: AVQI mean comparison between Praat and Matlab at T0.

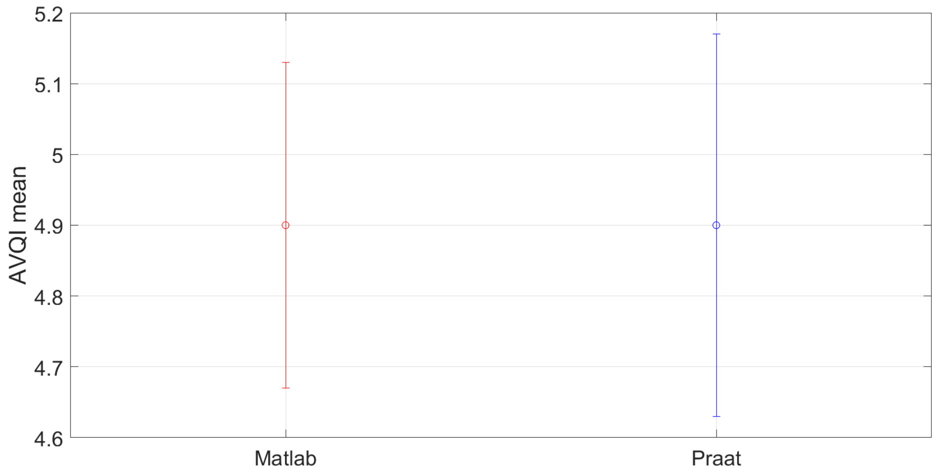

Figure 3.8: AVQI mean comparison between Praat and Matlab at T1.

	ID	CPPS(dB)	HNR(dB)	Shimmer(%)	Shimmer(dB)	Slope(dB)	Tilt (dB)	AVQI
In-clinic	01	11.7	6.7	13.5	1.3	-24.9	-7.5	5.2
	03	8.9	11.9	12.1	1.2	-29.2	-3.8	7.2
	05	12.9	12.1	11.0	1.0	-20.1	-8.2	3.8
	06	12.5	9.7	13.9	1.4	-18.4	-9.6	5.3
	08	11.1	9.7	12.1	1.1	-23.4	-9.3	5.4
	11	7.9	8.8	14.3	1.4	-28.9	-8.9	6.7
	13	10.3	11.4	12.1	1.2	-26.3	-6.8	5.8
	15	11.1	11.8	10.2	1.1	-23.5	-5.3	5.9
	19	12.0	9.3	9.5	1.0	-11.6	-7.6	5.1
	20	9.4	7.1	14.4	1.4	-27.7	-7.8	6.6
Telerehabilitation	02	12.0	13.0	9.3	1.0	-20.5	-7.3	4.8
	04	11.6	8.0	21.9	2.0	-26.4	-8.6	5.0
	07	11.1	8.1	23.4	2.3	-25.7	-8.4	6.9
	09	11.5	11.1	12.6	1.2	-19.9	-8.4	5.1
	10	12.4	11.8	11.1	1.0	-23.7	-8.5	4.1
	12	12.1	11.6	10.4	0.9	-19.1	-9.7	3.9
	14	6.5	6.7	18.5	1.7	-29.5	-7.7	8.4
	16	11.8	12.4	10.8	1.0	-17.4	-8.5	4.5
	17	12.7	10.0	50.5	4.8	-23.6	-6.2	6.3
	21	11.5	12.9	9.6	0.9	-21.8	-7.5	4.6

Table 3.5: Matlab results for the AVQI extraction at T0.

	ID	CPPS(dB)	HNR(dB)	Shimmer(%)	Shimmer(dB)	Slope(dB)	Tilt (dB)	AVQI
In-clinic	01	12.1	11.0	13.5	1.3	-25.5	-8.7	4.7
	03	10.9	12.3	8.9	0.9	-25.3	-4.9	6.2
	05	13.1	12.6	9.1	0.9	-17.5	-7.7	4.0
	06	12.6	10.7	10.7	1.1	-16.5	-10.5	3.8
	11	8.5	10.5	12.3	1.2	-27.3	-8.3	6.2
	13	9.9	12.2	11.9	1.1	-26.2	-7.3	5.7
	15	12.3	12.6	9.5	1.0	-17.4	-5.5	5.1
	19	11.6	9.9	11.2	1.1	-17.6	-7.6	5.0
	20	11.5	8.4	15.0	1.5	-25.2	-8.2	5.6
Telerehabilitation	02	11.9	14.7	8.5	0.9	-11.4	-6.7	5.1
	04	12.2	13.8	10.1	1.1	-26.7	-9.3	4.0
	07	12.7	11.0	29.1	2.8	-24.9	-8.7	4.8
	08	12.5	13.7	9.9	1.0	-14.6	-9.5	4.0
	09	10.5	9.7	12.6	1.2	-19.6	-8.4	5.7
	10	13.4	13.3	7.2	0.9	-12.5	-8.3	4.0
	12	13.0	13.1	8.6	0.9	-12.1	-8.7	3.1
	14	8.8	9.0	14.7	1.4	-25.8	-7.0	7.1
	16	11.2	12.3	10.7	1.1	-15.5	-9.1	5.0
	17	11.7	8.5	34.9	3.5	-22.3	-6.8	6.3
	21	11.9	13.8	8.7	0.8	-20.4	-8.5	4.1

Table 3.6: Matlab results for the AVQI extraction at T1.

	ID	CPPS(dB)	HNR(dB)	Shimmer(%)	Shimmer(dB)	Slope(dB)	Tilt (dB)	AVQI
In-clinic	01	10.0	11.0	13.5	1.2	-24.8	-7.1	6.0
	03	7.9	15.3	13.1	1.2	-29.2	-3.6	7.8
	05	12.6	12.5	10.1	0.9	-20.1	-8.0	4.1
	06	9.9	8.6	12.6	1.1	-18.4	-9.3	5.6
	08	9.8	9.9	14.8	1.3	-23.3	-9.0	5.7
	11	7.1	8.7	15.9	1.4	-28.8	-8.7	7.1
	13	9.4	11.2	14.2	1.3	-26.3	-6.6	6.4
	15	10.0	12.9	10.9	1.1	-23.4	-4.8	6.4
	19	12.9	9.4	9.8	1.0	-11.7	-7.3	4.7
	20	9.3	10.8	12.4	1.1	-27.7	-7.5	6.0
Telerehabilitation	02	11.7	12.8	10.2	1.0	-20.5	-6.7	5.1
	04	10.7	11.9	11.0	1.0	-26.3	-8.3	4.9
	07	10.1	9.6	13.0	1.1	-25.7	-8.2	5.5
	09	11.0	10.4	10.3	0.9	-19.9	-8.1	4.9
	10	11.9	11.4	11.6	1.0	-23.7	-8.3	4.9
	12	12.7	11.5	11.3	1.1	-19.1	-9.4	3.9
	14	5.8	5.6	18.6	1.6	-29.6	-7.6	8.3
	16	11.8	11.5	10.9	1.0	-17.4	-8.2	4.6
	17	10.0	9.0	10.2	0.9	-23.5	-6.0	5.7
	21	11.2	12.3	9.5	0.9	-21.8	-7.2	4.9

Table 3.7: Praat results for the AVQI extraction at T0: *'Version 2'* of the elaboration.

	ID	CPPS(dB)	HNR(dB)	Shimmer(%)	Shimmer(dB)	Slope(dB)	Tilt (dB)	AVQI
In-clinic	01	11.4	12.3	11.1	1.0	-25.5	-8.5	4.5
	03	9.2	13.9	10.1	1.0	-25.3	-4.6	6.7
	05	13.7	13.1	6.5	0.7	-17.6	-7.3	3.4
	06	11.7	10.3	10.5	1.0	-16.5	-10.2	4.4
	08	12.2	13.2	10.7	1.0	-15.0	-9.1	4.4
	11	7.6	10.2	12.8	1.1	-27.3	-8.2	6.7
	13	9.8	12.1	12.8	1.2	-26.1	-7.0	6.1
	15	12.1	12.7	8.2	0.9	-17.4	-4.9	5.2
	19	12.4	9.4	12.3	1.1	-17.6	-7.3	4.7
	20	11.4	11.6	10.1	1.0	-25.1	-7.9	4.7
Telerehabilitation	02	13.2	14.9	9.6	0.9	-11.4	-6.0	4.6
	04	11.1	13.7	8.3	0.9	-26.8	-8.8	4.4
	07	12.0	11.4	10.7	0.9	-24.8	-8.4	4.1
	09	10.2	9.3	11.3	1.0	-19.5	-8.2	5.4
	10	13.7	12.5	7.2	0.9	-12.4	-7.7	4.0
	12	15.5	14.4	6.5	0.7	-12.0	-8.4	2.5
	14	7.6	8.3	16.2	1.5	-25.7	-6.8	7.7
	16	12.3	11.7	8.5	0.8	-15.4	-8.9	4.1
	17	10.5	8.3	13.6	1.1	-22.2	-6.5	5.8
	21	12.6	13.2	8.6	0.8	-20.4	-8.2	3.8

Table 3.8: Praat results for the AVQI extraction at T1: *'Version 2'* of the elaboration.

3.2 Software analysis to compute SPL

The comparison between the software applications is conducted in terms of Sound Pressure Level (SPL) in this section. This analysis allows to have an immediate idea of the intensity of the voice, which is a parameter that is altered in multiple sclerosis subjects.
The SPL, introduced in 2.4, is extracted from Praat and Matlab and it involves in its formula the use of the Root Mean Suqare (RMS). VOXplot does not provide the RMS among all the available values, so the computation of this parameter is impossible using this software.
This parameter is extracted considering for each subject a monologue recording that lasts about one minute.
The extraction is conducted both on the recordings at time T0 and T1, in order to subsequently evaluate a possible improvement in the intensity of the voice after the therapy as well as the non-inferiority of the telerehabilitation therapy compared to the in-clinic one, as will be seen in 3.4.3.
The SPL values reported in the following tables are refferred to a distance of 30cm from the mouth, as it is the distance at which the in-air microphone is normally placed. The results are expressed in terms of SPLmean $\pm$ one standard deviation, where the average is made between all the patients without a distinction, in this first phase of the analysis, between the in-clinic and the telerehabilitation group.
Table 3.9 shows the SPL values extracted from Praat at T0 and T1 respectively. At T0 Praat results in an SPL average value of (71.5$\pm$0.56)dB; at T1 the SPL average value is (72.9$\pm$0.45)dB.
Table 3.10 shows the SPL values extracted from Matlab at T0 and T1 respectively. At T0 Matlab results in an SPL average value of (71.1$\pm$0.59)dB; at T1 the SPL average value is (72.8$\pm$0.44)dB.
The comparison between the results of the two software applications is shown in Fig. 3.9 and in Fig. 3.10, respectively at time T0 and T1. Instead, in Fig. 3.11 and in Fig. 3.12 the comparison between the mean SPL values and the uncertainty bars related to the software application are rapresented, for T0 and T1 respectively. It is evident that the results returned by the two software are very similar to each other, both at time T0 and T1, confirming the validity of both software applications in the extraction of the SPL.
However, it should be underlined that Matlab allows to conduct a more in-depth analysis of the SPL by returning the vector of RMS values from which various statistics can be extracted.
Instead, Praat only returns the SPL average value not allowing a statistical analysis.

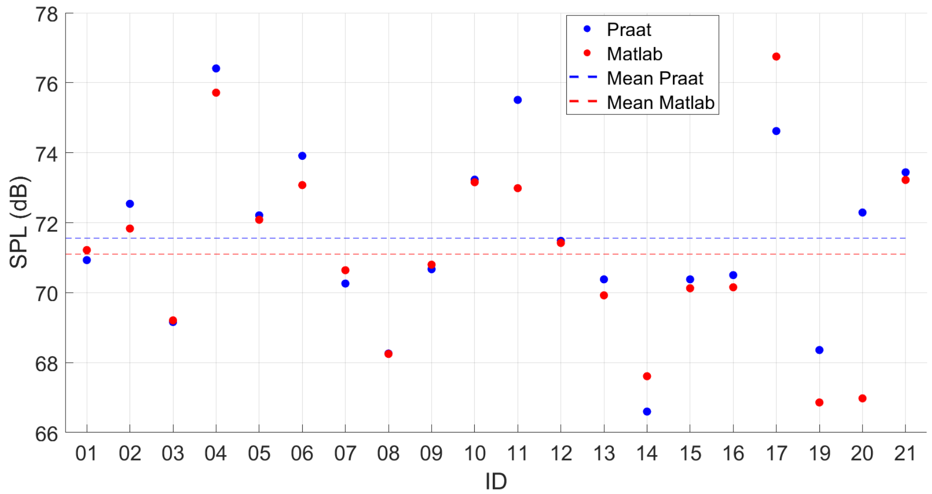

Figure 3.9: SPL comparison between Praat and Matlab at T0.

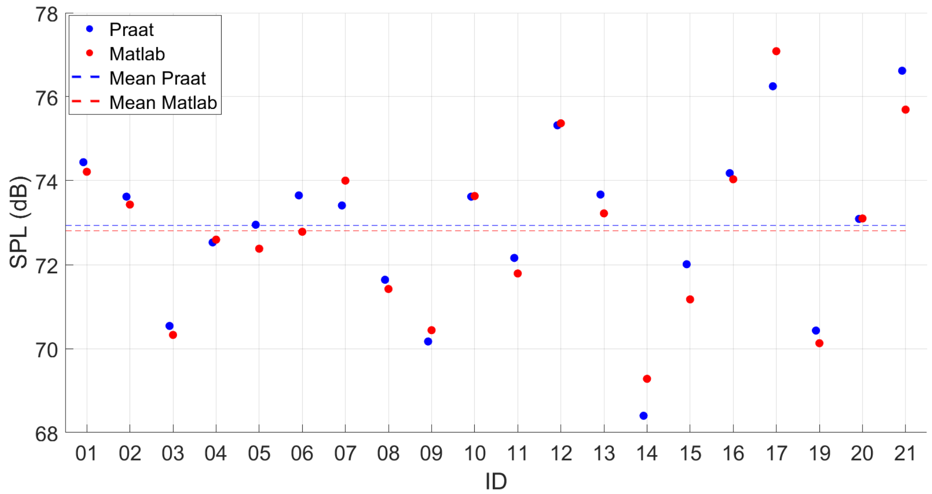

Figure 3.10: SPL comparison between Praat and Matlab at T1.

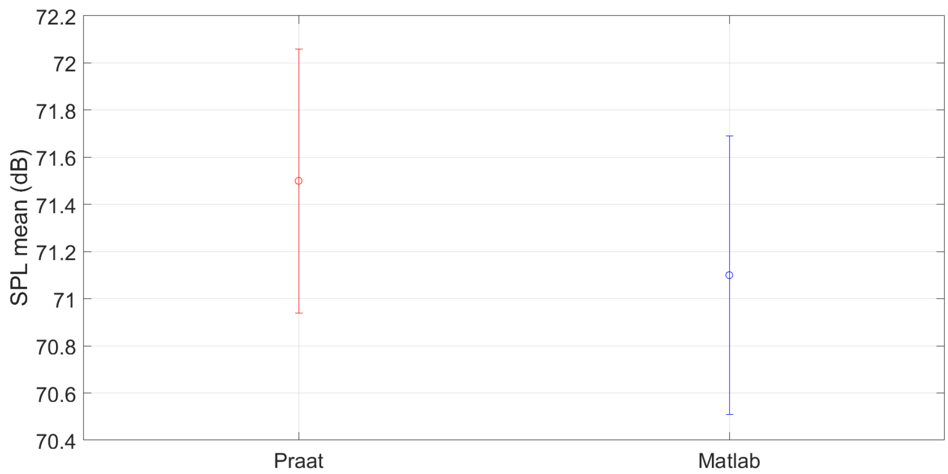

Figure 3.11: SPL mean comparison between Praat and Matlab at T0.

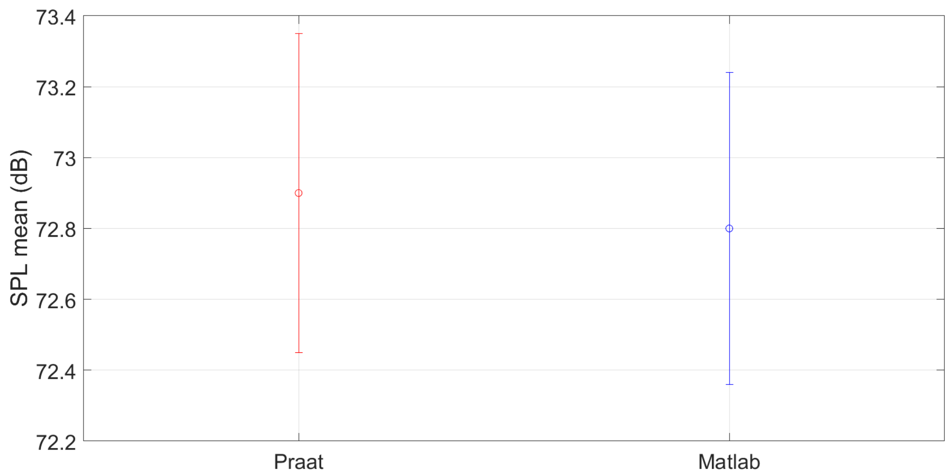

Figure 3.12: SPL mean comparison between Praat and Matlab at T1.

	ID	SPL(dB) at T0	SPL(dB) at T1
In-clinic	01	70.9	74.4
	03	69.1	70.5
	05	72.2	72.9
	06	73.9	73.6
	08	68.2	71.6
	11	75.5	72.1
	13	70.3	73.6
	15	70.3	72.0
	19	68.3	70.4
	20	72.2	73.0
Telerehabilitation	02	72.5	73.6
	04	76.4	72.5
	07	70.2	73.4
	09	70.6	70.1
	10	73.2	73.6
	12	71.4	75.3
	14	66.6	68.4
	16	70.5	74.1
	17	74.6	76.2
	21	73.4	76.6

Table 3.9: Praat results for SPL extraction at T0 and T1.

	ID	SPL(dB) at T0	SPL(dB) at T1
In-clinic	01	71.2	74.2
	03	69.2	70.3
	05	72.0	72.3
	06	73.0	72.7
	08	68.2	71.4
	11	72.9	71.7
	13	69.9	73.2
	15	70.1	71.1
	19	66.8	70.1
	20	66.9	73.1
Telerehabilitation	02	71.8	73.4
	04	75.7	72.6
	07	70.6	74.0
	09	70.8	70.4
	10	73.1	73.6
	12	71.4	75.3
	14	67.6	69.2
	16	70.1	74.0
	17	76.7	77.0
	21	73.2	75.6

Table 3.10: Matlab results for SPL extraction at T0 and T1.

3.3 Microphone comparison to compute WS

All the recordings of the Multiple Sclerosis (MS) dataset are acquired both from a vocal recorder, equipped with a microphone in air, and from a contact microphone (Vocal Holter, VH). Therefore, an analysis of the results in terms of Warning Score (WS) is conducted comparing the two acquisition systems.

The WS, introduced in 2.5.2, is computed on a sustained vowel file considering all the three recordings available. The parameters of interest are the Local Jitter (%), the Local Shimmer (%), the CPPSmedian (Cepstral Peak Prominence Smoothed median, dB) and the CPPSstd (Cepstral Peak Prominence Smoothed standard deviation, dB).

Table 3.11 and Table 3.12 show the WS values and the parameters involved in its calculation extracted from the microphone in air at T0 and T1 respectively.

Table 3.13 and Table 3.14 show the WS values and the parameters involved in its calculation extracted from the Vocal Holter at T0 and T1 respectively.

The comparison between the results of the two microphones is shown in Fig. 3.13 and in Fig. 3.14, respectively at T0 and T1.

From the results, it is evident that the contact microphone returns values of WS usually lower than those returned by the microphone in air, both at time T0 and T1.

This is caused by the fact that the VH returns Local Shimmer values lower than those returned by the microphone in air, often below the threshold that identifies a patological voice, obtaining a score equal to -1 (healthy voice) and not +1 (patological voice).

Infact, the Local Shimmer mean value for the contact microphone at T0 is equal to 4.0±0.59, while for the microphone in air is equal to 8.0±0.71.

This could be due to the fact that the VH is placed on the neck and not in front of the mouth, so it is less sensitive to the presence of noice. Instead, the microphone in air acquires also the environmental noice, which causes random fluctuations in amplitude and, consequently, the Local Shimmer values become higher.

However, this does not happen for the other parameters involved in the WS computation, which appear to be very similar between the two acquisition systems. Infact, considering as example the time T0, the Local Jitter mean is equal to 0.87±0.13 for the VH, while for the microphone in air is equal to 0.83±0.10; the CPPSmedian mean is equal 15.7±0.45 to for the VH, while for the microphone in air is equal to 14.3±0.51; the CPPSstd mean is equal to 1.85±0.17 for the VH, while for the microphone in air is equal to 2.01±0.11.

For the previous reasons, the contact microphone is considered to be more precise in extracting the Local Shimmer value and so, it is believed to return more reliable values of WS.

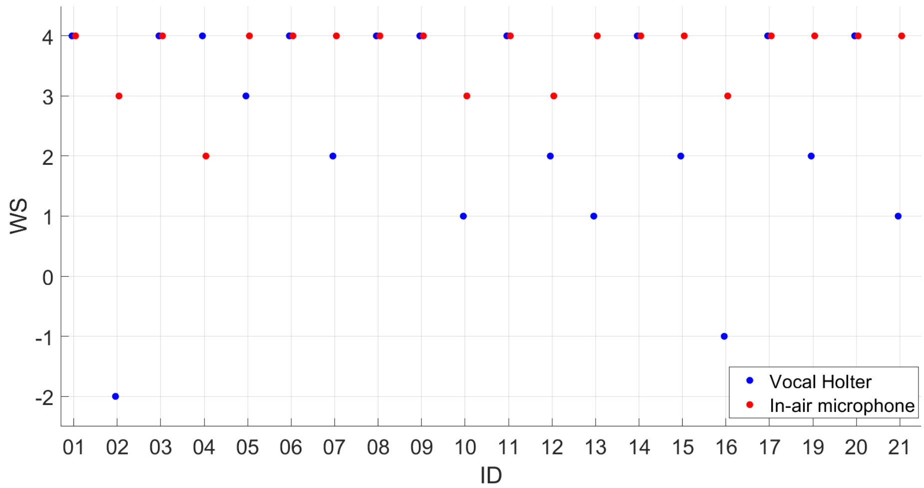

Figure 3.13: Microphone comparison in terms of WS at T0.

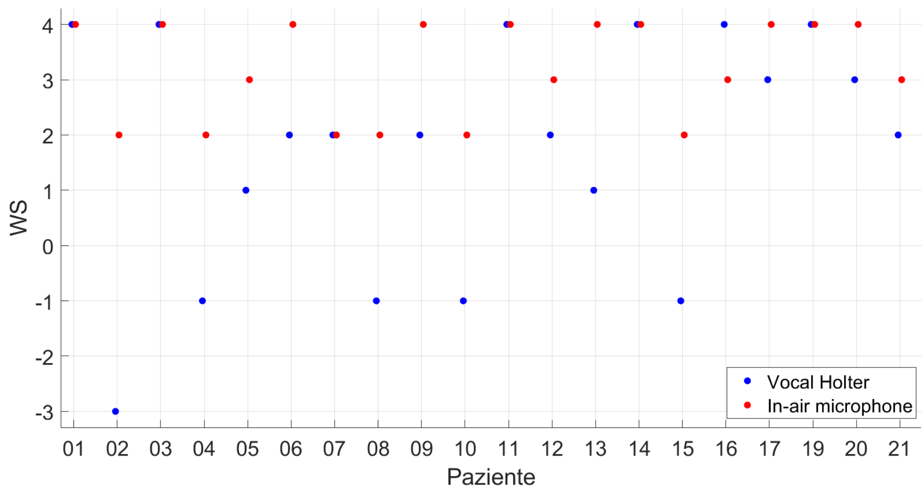

Figure 3.14: Microphone comparison in terms of WS at T1.

	ID	Jitter(%)	Shimmer(%)	CPPSmedian(dB)	CPPSstd(dB)	WS
In-clinic	01	0.82	6.88	14.3	2.3	+4
	03	1.67	7.81	11.2	2.6	+4
	05	0.63	6.42	17.1	1.5	+4
	06	0.78	8.80	17.5	2.1	+4
	08	0.67	9.91	13.9	2.1	+4
	11	1.92	11.32	10.2	2.0	+4
	13	0.72	8.95	13.5	2.0	+4
	15	0.60	5.61	15.0	1.9	+4
	19	0.64	6.81	14.7	1.7	+4
	20	1.22	8.34	13.1	2.7	+4
Telerehabilitation	02	0.42	4.98	15.3	1.5	+3
	04	0.29	18.86	16.4	3.4	+2
	07	1.35	11.95	14.9	1.5	+4
	09	0.84	6.78	14.9	1.9	+4
	10	0.36	5.89	16.0	1.7	+3
	12	0.38	5.26	15.8	1.5	+3
	14	1.77	8.05	8.0	2.2	+4
	16	0.37	5.00	14.7	1.4	+3
	17	0.65	7.55	15.2	1.7	+4
	21	0.53	6.29	13.8	1.4	+4

Table 3.11: Microphone in air results for the WS extraction at T0.

	ID	Jitter(%)	Shimmer(%)	CPPSmedian(dB)	CPPSstd(dB)	WS
In-clinic	01	0.50	5.75	14.5	1.8	+4
	03	0.83	5.31	15.1	2.1	+4
	05	0.38	4.16	16.9	1.3	+3
	06	0.44	6.85	17.3	1.3	+4
	08	0.23	5.01	16.4	1.4	+2
	11	1.26	8.45	11.7	1.9	+4
	13	0.71	9.22	13.2	1.7	+4
	15	0.27	5.27	15.7	1.4	+2
	19	0.51	6.56	14.0	1.9	+4
	20	1.50	11.96	15.9	2.4	+4
Telerehabilitation	02	0.21	5.00	14.4	1.5	+2
	04	0.23	4.60	17.3	1.4	+2
	07	0.28	6.68	16.5	1.7	+2
	09	1.59	8.26	12.8	3.0	+4
	10	0.20	4.61	17.5	1.3	+2
	12	0.32	3.67	15.8	1.5	+3
	14	0.79	12.33	11.0	2.1	+4
	16	0.34	4.97	14.5	1.8	+3
	17	0.49	8.87	16.2	1.7	+4
	21	0.35	4.23	14.2	1.3	+3

Table 3.12: Microphone in air results for the WS extraction at T1.

	ID	Jitter(%)	Shimmer(%)	CPPSmedian(dB)	CPPSstd(dB)	WS
In-clinic	01	0.78	2.62	16.6	2.7	+4
	03	1.61	6.03	10.8	1.9	+4
	05	0.65	2.55	16.8	1.5	+3
	06	1.28	5.82	17.6	3.2	+4
	08	1.05	3.44	13.4	1.9	+4
	11	1.38	8.53	13.0	1.7	+4
	13	0.70	2.07	15.4	1.2	+1
	15	0.51	1.51	15.6	1.3	+2
	19	0.47	2.18	17.2	1.4	+2
	20	2.72	6.69	15.5	3.9	+4
Telerehabilitation	02	0.23	1.10	18.0	1.0	-2
	04	0.45	3.66	16.7	1.4	+4
	07	0.42	5.05	15.3	1.1	+2
	09	1.17	4.73	16.9	1.8	+4
	10	0.33	2.38	18.7	1.6	+1
	12	0.64	1.90	16.9	1.7	+2
	14	1.69	11.59	12.0	3.1	+4
	16	0.26	1.89	15.0	1.1	-1
	17	0.64	5.19	17.1	1.6	+4
	21	0.40	2.40	15.3	1.0	+1

Table 3.13: VH results for the WS extraction at T0.

	ID	Jitter(%)	Shimmer(%)	CPPSmedian(dB)	CPPSstd(dB)	WS
In-clinic	01	0.59	3.13	17.4	1.8	+4
	03	0.55	3.56	14.8	1.8	+4
	05	0.38	1.61	17.8	1.7	+1
	06	0.56	4.02	18.6	1.2	+2
	08	0.20	1.57	17.7	0.9	-1
	11	0.99	5.13	13.4	1.5	+4
	13	0.48	2.07	15.5	1.0	+1
	15	0.26	1.23	17.2	0.9	-1
	19	0.82	2.61	15.5	1.4	+4
	20	1.01	3.51	18.4	2.6	+3
Telerehabilitation	02	0.18	1.09	18.4	0.8	-3
	04	0.22	1.86	18.6	1.3	-1
	07	0.25	11.55	17.2	1.3	+2
	09	1.18	10.00	17.4	0.8	+2
	10	0.21	1.74	18.6	1.5	-1
	12	0.32	8.04	18.3	1.9	+2
	14	1.12	6.37	13.8	1.9	+4
	16	1.09	3.27	15.8	1.7	+4
	17	0.46	4.35	18.7	1.8	+3
	21	0.40	2.72	15.9	1.1	+2

Table 3.14: VH results for the WS extraction at T1.

3.4 Evaluation of the LSVT-LOUD theraphy

The evaluation of the effectiveness of the LSVT-LOUD therapy as well as the non-inferiority of the telerehabilitation treatment compared to the in-clinic one is conducted in this section. This is done in terms of vocal indexes (AVQI and WS) and in terms of vocal parameters.

3.4.1 Evaluation in terms of AVQI

The evaluation of the effectiveness of the LSVT-LOUD therapy is firstly conducted. In this first part of the analysis, the distinction between the in-clinic and the telerehabilitation group is not made, so the results are referred to all the patients. This is done in order to evaluate the improvement of the patients after the therapy, in terms of ΔAVQI.
The ΔAVQI is computed as the difference between the AVQI values at time T0 and T1: a positive value of this difference refers to an improvement of the vocal situation, while a negative values reveals a worsening.
This is performed for all the software applications and the results are shown in Fig. 3.15 for VOXplot, in Fig. 3.16 for Praat and in Fig. 3.17 for Matlab, where the ΔAVQI values and a line at zero corresponding to the pathology improvement threshold are represented. Values over this threshold indicate an improvement, while values below reveal a worsening.

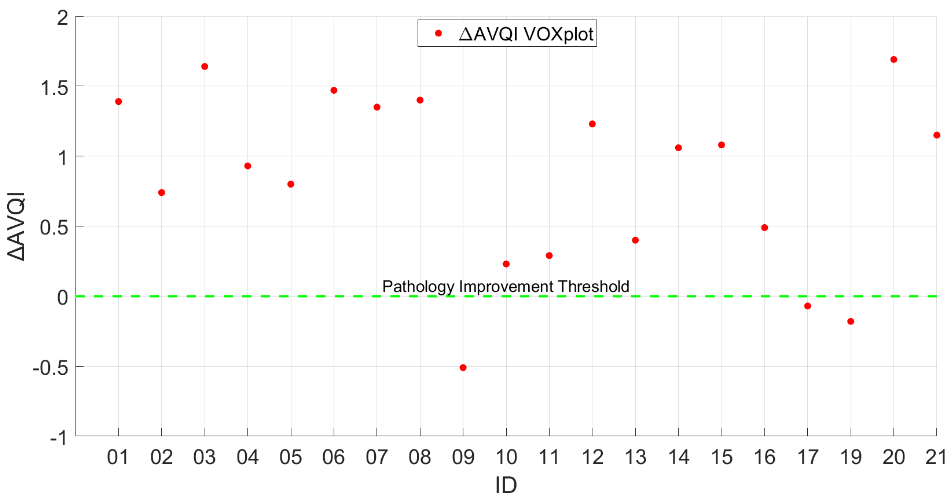

Figure 3.15: ΔAVQI values in VOXplot.

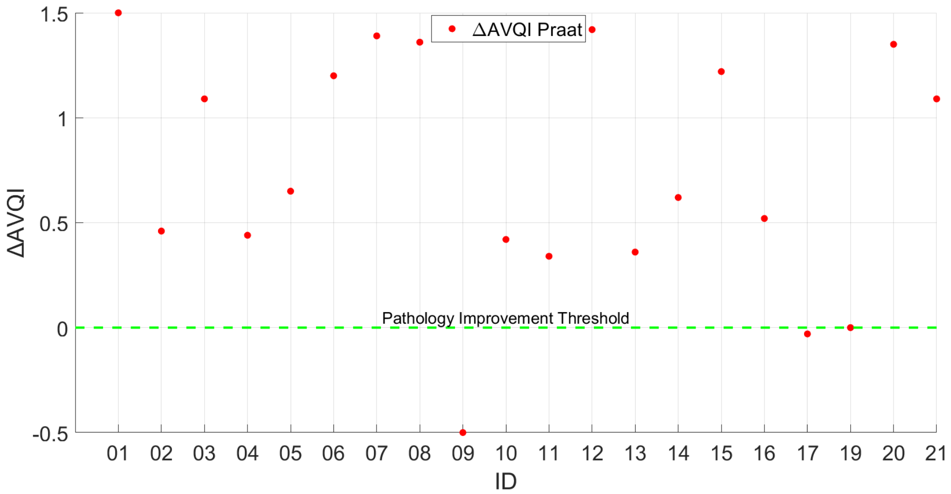

Figure 3.16: ΔAVQI values in Praat.

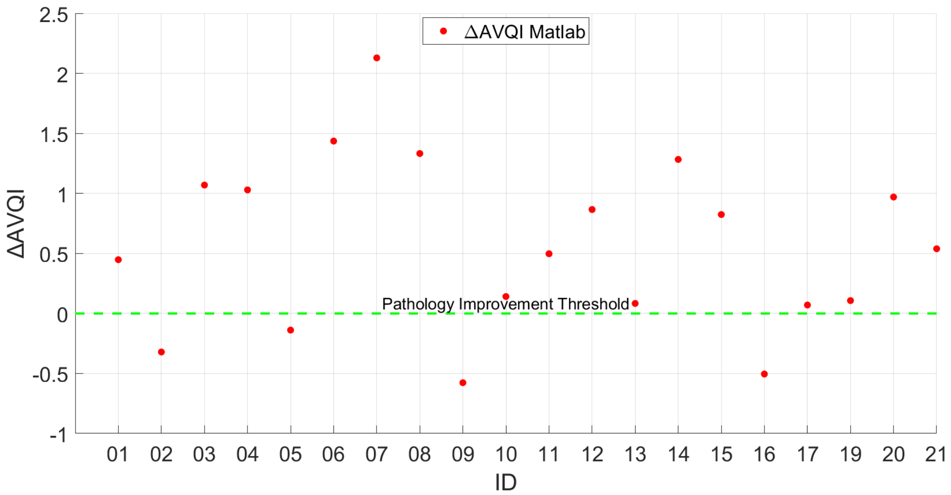

Figure 3.17: ΔAVQI values in Matlab.

For all the software applications, the analysis results for most patient in a positive value of ΔAVQI, that indicates an improvement of the vocal situation after the therapy, confirming the effectiveness of the LSVT-LOUD therapy.

At this point, the distinction between the in-clinic and the telerehabilitation group is made, with the aim to understand whether the improvement previously found occurs equally in both groups of patients. This can lead to the assessment of the non-inferiority of the telerehabilitation treatment compared to the in-clinic one.
The results are presented in terms of ΔAVQI avereged distinctly for the two groups of subjects. The uncertainty bars added to the averaged ΔAVQI values are representative of both the variability of the voice of the subject during the esecution of the sustained vowel and the type of therapy followed, having taken into account all three vocal recordings.
VOXplot results in an average value of ΔAVQI equal to 1.0±0.19 for the in-clinic group and 0.6±0.18 for the telerehabilitation group.
Praat results in an average value of ΔAVQI equal to 0.9±0.15 for the in-clinic group and 0.6±0.20 for the telerehabilitation group.
Matlab results in an average value of ΔAVQI equal to 0.6±0.17 for the in-clinic group and 0.4±0.24 for the telerehabilitation group.
These results are shown in Fig. 3.18 for VOXplot, in Fig. 3.19 for Praat and in Fig. 3.20 for Matlab.

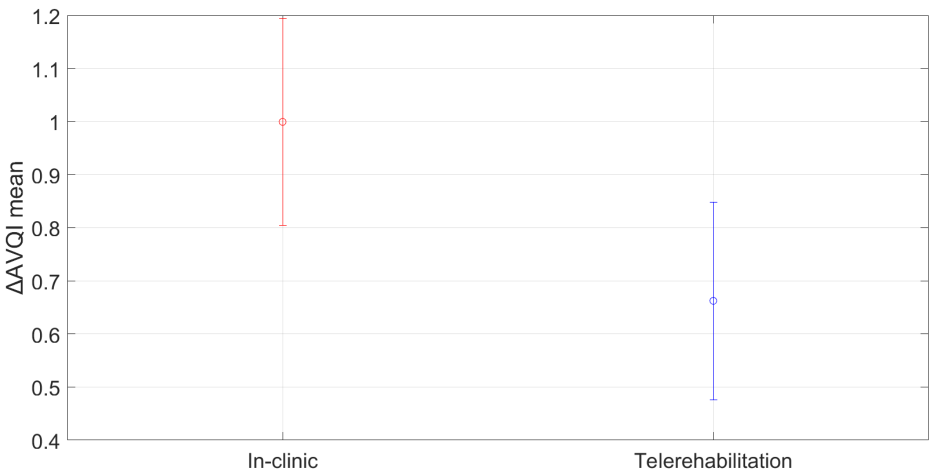

Figure 3.18: Comparison of therapies for VOXplot in terms of ΔAVQI.

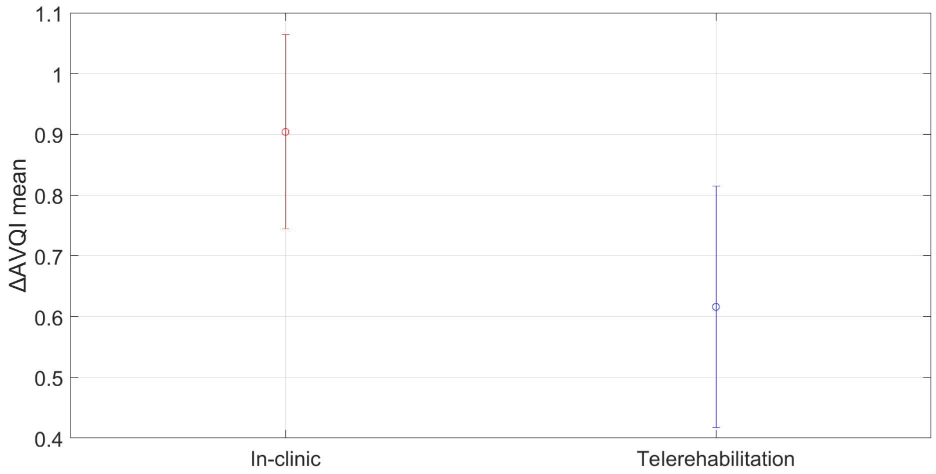

Figure 3.19: Comparison of therapies for Praat in terms of ΔAVQI.

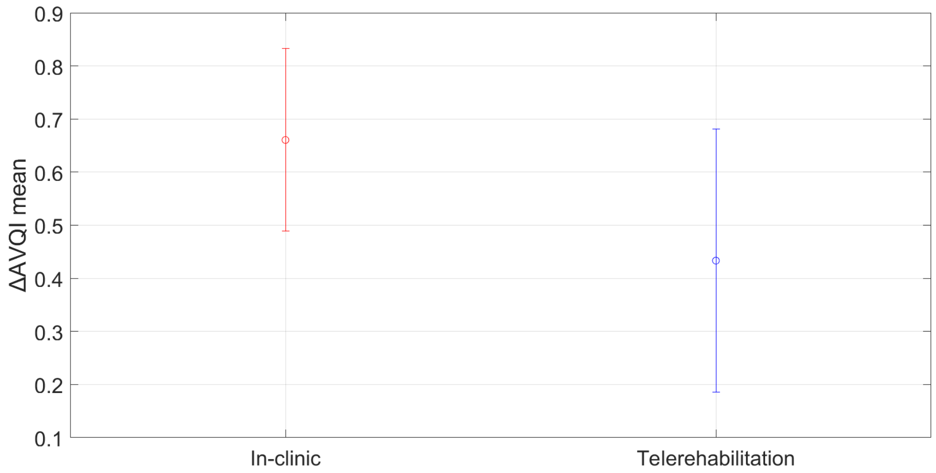

Figure 3.20: Comparison of therapies for Matlab in terms of ΔAVQI.

For all the software applications, it is further evident that both the in-clinic and the telerehabilitation therapy produce a vocal improvement, returning values of avereged ΔAVQI above zero.
In addition, the improvements produced following the two therapies are comparable, returning similar values of avereged ΔAVQI, as well as overlapping bands of uncertainty.
This confirms the non-inferiority of the telerehabilitation treatment compared to the in-clinic one, so the vocal improvement produced by the two therapies is equivalent.

3.4.2 Evaluation in terms of WS

The evaluation of the non-inferiority of the LSVT-LOUD treatment administred in telerehabilitation is conducted in this subsection in terms of WS. In particular, the ΔWS is employed and it is calculated as the difference between the WS values at T0 and T1.
The results extracted from the contact microphone are analyzed, being considered more reliable as reported in 3.3. The values of ΔWS at T0 and T1 are rapresented for each patient: if the value at is higher than zero, a vocal improvement is present. These results are shown in Fig. 3.21 for the in-clinic group and in Fig. 3.22 for the telerehabilitation group.

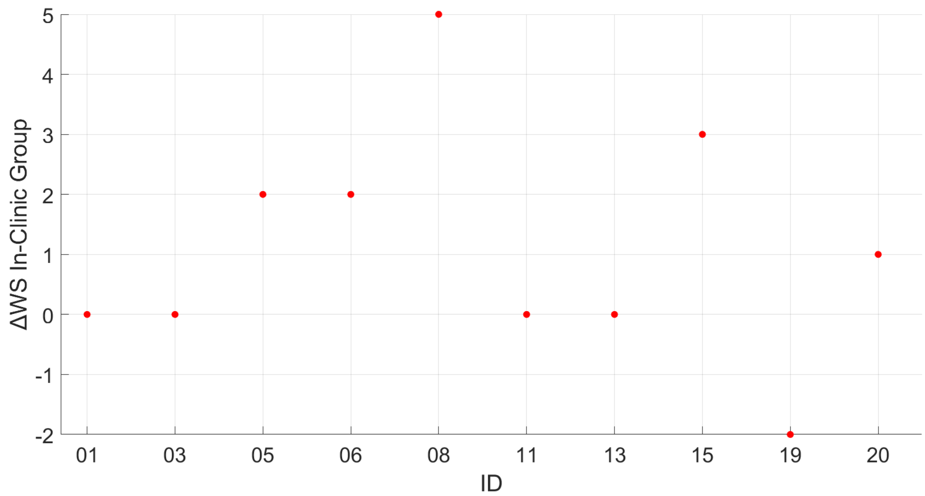

Figure 3.21: Evaluation of the improvement in terms of ΔWS for the in-clinic group.

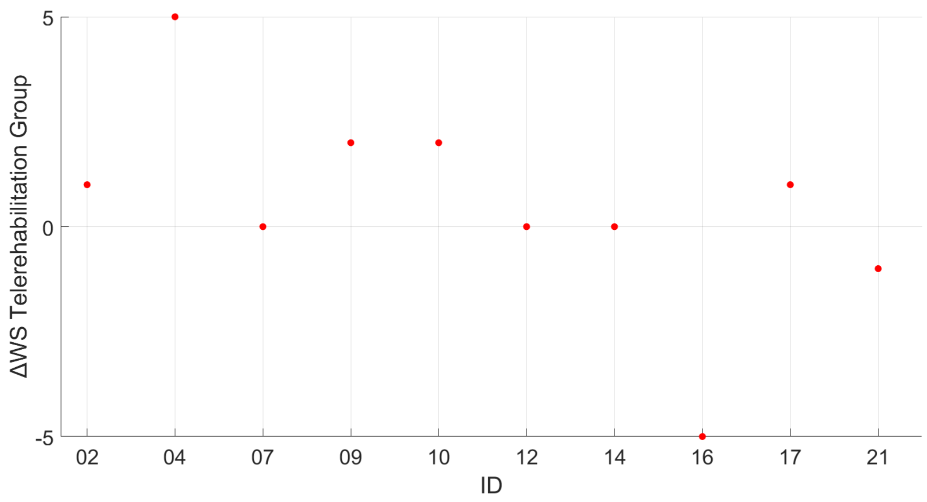

Figure 3.22: Evaluation of the improvement in terms of ΔWS for the telerehabilitation group.

From the results it is evident that both therapies are effective, as the improvement is present for the same number of patients for the in-clinic group and for the telerehabilitation group. Therefore, the Tele-LSVT-LOUD therapy is not inferior to the in-clinic one, also in terms of WS.
At the same time, the results evidence that for a few subjects of both groups, there is a worsening of the WS after the therapy. This may be due not to the therapy itself, but to the way in which the task is performed: performing the vowel with less effort can result in a worsening of the parameters of interest and of the WS.

3.4.3 Evaluation in terms of vocal parameters

The evaluation of the non-inferiority of the LSVT-LOUD treatment administred in telerehabilitation is conducted in this subsection in terms of vocal parameters extracted from both the monologue and the sustained vowel recording.
From the monologue recording the SPL is extracted, as it allows to have an immediate idea of the intensity of the voice and so of a possible improvement after the therapy.
The SPL is not a significant parameter for the sustained vowel recording, so from this recording the Local Jitter, Local Shimmer, HNR and CPPS are extracted and they are shown in Table 3.15 and Table 3.16, respectively at T0 and T1.
The results, extracted from Praat and expressed in terms of Δ values (difference between values at different acquisition times), are analyzed. A positive Δ value

refers to an improvement of the voice, while e nagative value reveals a worsening. Considering firstly the monologue, the ΔSPL is computed as the difference between the SPL values at time T1 and T0.
The results are shown in Fig. 3.23, where the ΔSPL values and a line at zero corresponding to the voice improvement threshold are represented, without distinction between the two groups of subjects. Values over this threshold indicate an improvement, while values below reveal a worsening.
The results are then expressed in terms of ΔSPL avereged for the two groups of subjects and the uncertainty bars, rapresentative of the therapy followed, are added to the averaged values, as shown in Fig. 3.24.
Praat results in an average value of ΔSPL equal to (1.3±0.65)dB for the in-clinic group and equal to (1.4±0.74)dB for the telerehabilitation group.
From the results, it is evident that both the in-clinic and the telerehabilitation therapy produce an improvement of the intensity of the voice, returning values of avereged ΔSPL above zero.
In addition, the improvements produced following the two therapies are comparable, returning similar values of avereged ΔSPL, as well as overlapping bands of uncertainty.
This confirms, also in the case of the SPL, the non-inferiority of the telerehabilitation treatment compared to the in-clinic one.

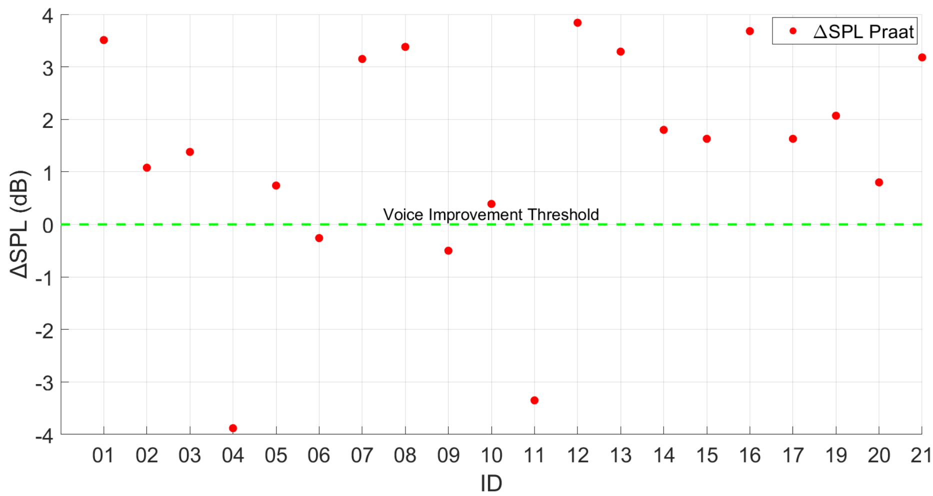

Figure 3.23: ΔSPL values in Praat.

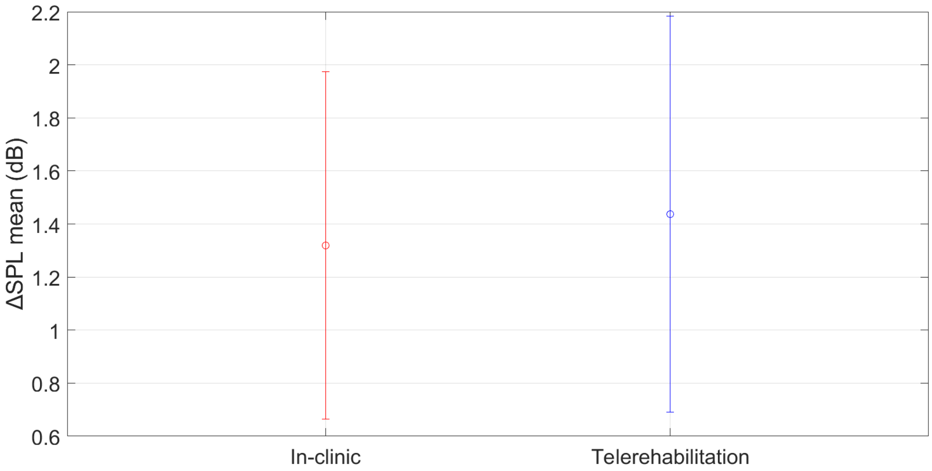

Figure 3.24: Comparison of therapies for Praat in terms of ΔSPL.

Considering the sustained vowel recording, the ΔJitter and ΔShimmer are computed as the difference between the values at time T0 and T1, while the ΔCPPS and ΔHNR are computed as the difference between T1 and T0. The results, expressed in terms of avereged Δ values for the two groups of subjects, are the following:

- ΔShimmer: (2.3±0.59)% for the in-clinic group and (1.4±0.86)% for the telerehabilitation group
- ΔJitter: (0.27±0.09)% for the in-clinic group and to (0.16±0.10)% for the telerehabilitation group
- ΔHNR: (1.9±0.82)dB for the in-clinic group and to (1.2±0.93)dB for the telerehabilitation group
- ΔCPPS: (1.9±0.66)dB for the in-clinic group and to (2.2±0.62)dB for the telerehabilitation group

They are shown in Fig. 3.25 for the ΔShimmer, in Fig. 3.26 for the ΔJitter, in Fig. 3.27 for the ΔHNR and in Fig. 3.28 for the ΔCPPS.
From these results, it is evident that both the in-clinic and the telerehabilitation therapy produce an improvement of the intensity of the voice, returning values above zero. In addition, the improvements produced following the two therapies are comparable, returning similar values of avereged Δ, as well as overlapping bands of uncertainty.

	ID	Jitter(%)	Shimmer(%)	HNR(dB)	CPPS(dB)
In-clinic	01	0.7	12.3	11.4	10.6
	03	1.8	7.3	16.4	7.9
	05	0.6	7.9	13.9	14.5
	06	1.1	8.8	11.9	11.0
	08	0.9	9.0	8.9	10.1
	11	1.2	13.1	8.1	7.3
	13	1.0	15.3	9.2	9.7
	15	0.8	7.5	14.6	10.9
	19	0.5	7.6	10.1	16.9
	20	1.0	10.1	12.7	11.9
Telerehabilitation	02	0.5	8.6	14.2	13.9
	04	0.5	6.8	12.9	12.5
	07	0.5	12.6	12.3	11.3
	09	0.8	9.3	13.8	12.4
	10	0.4	8.4	13.8	13.7
	12	0.4	6.8	14.6	15.7
	14	1.7	8.6	9.1	5.9
	16	0.5	9.6	16.1	13.5
	17	0.8	6.1	11.3	11.9
	21	0.6	7.6	13.8	12.2

Table 3.15: Vocal parameters extracted in Praat from the sustained vowel at T0.

	ID	Jitter(%)	Shimmer(%)	HNR(dB)	CPPS(dB)
In-clinic	01	0.5	8.3	15.0	12.9
	03	1.1	5.8	15.6	10.5
	05	0.5	4.1	16.1	15.6
	06	0.6	6.2	11.8	13.9
	08	0.3	6.0	17.1	15.7
	11	1.4	9.0	11.6	8.0
	13	0.7	15.1	10.6	10.2
	15	0.3	5.1	14.8	14.5
	19	0.7	9.2	11.6	14.4
	20	0.6	7.6	12.6	14.7
Telerehabilitation	02	0.3	5.1	17.2	17.1
	04	0.3	4.4	19.1	15.1
	07	0.3	8.8	14.0	12.6
	09	1.1	8.8	10.0	10.4
	10	0.2	4.3	17.6	19.3
	12	0.4	3.0	17.5	20.1
	14	0.7	11.7	8.3	8.8
	16	0.5	7.0	14.3	15.4
	17	0.5	8.9	10.2	13.2
	21	0.6	5.6	16.3	13.8

Table 3.16: Vocal parameters extracted in Praat from the sustained vowel at T1.

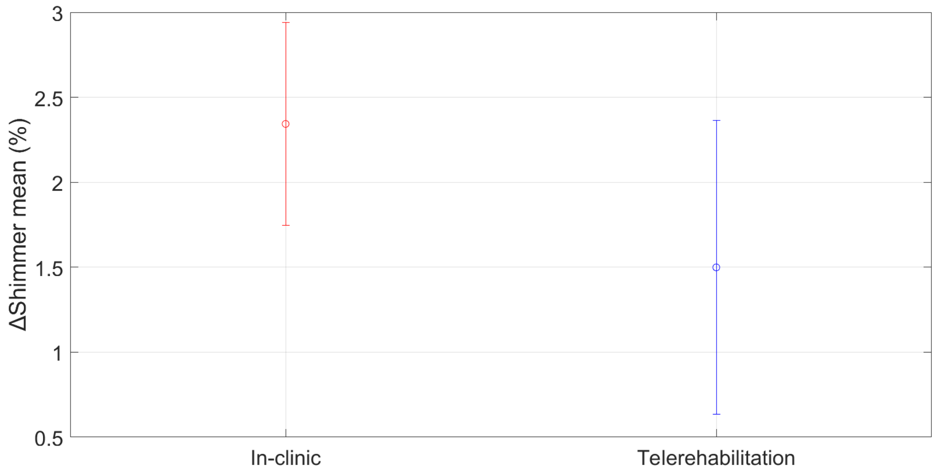

Figure 3.25: Comparison of therapies for Praat in terms of ΔShimmer.

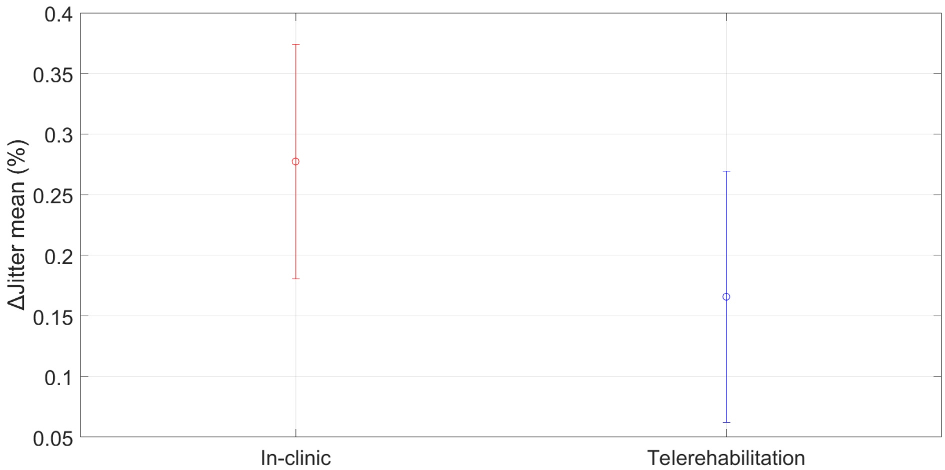

Figure 3.26: Comparison of therapies for Praat in terms of ΔJitter.

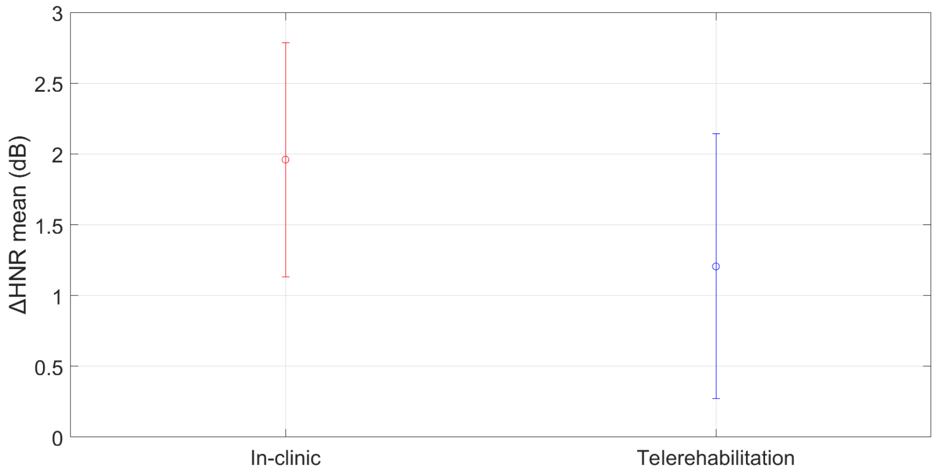

Figure 3.27: Comparison of therapies for Praat in terms of ΔHNR.

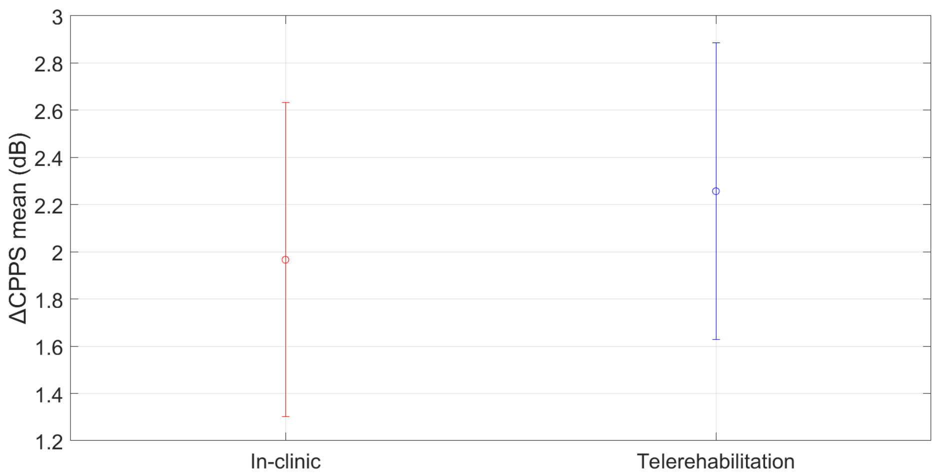

Figure 3.28: Comparison of therapies for Praat in terms of ΔCPPS.

3.5 VHI results

The Voice Handicap Index (VHI) is employed to understand the perception of the subjects of their voice before the therapy (at T0) and after the therapy (at T1), evaluating their vocal handicap.
The VHI results are shown in Table 3.17 for the in-clinic group and in Table 3.18 for the telerehabilitation group.
An higher VHI score is correlated with a worse perception of the disease, while a lower score is correlated with a better perception of the disease. Therefore, the lower the VHI, the better.

ID	VHI T0	VHI T1
01	53	26
03	82	85
05	74	67
06	55	10
08	57	49
11	78	64
13	30	25
15	18	15
19	34	41
20	21	21

Table 3.17: VHI results for in-clinic group.

ID	VHI T0	VHI T1
02	69	62
04	19	19
07	22	7
09	58	36
10	24	11
12	16	12
14	38	47
16	13	9
17	33	18
21	37	15

Table 3.18: VHI results for telerehabilitation group.

The results reveal that the VHI scores at T1 are lower than those obtained at T0 for most patients: this means that the subjects perceive an improved vocal situation after the therapy.
This happens for all the patients except for the subjects with ID 04 and 20 which have the same VHI score at T0 and T1, not perceiving any improvement, and for the subjects with ID 14 and 19 that have a higher VHI score at T1 than at T0, perceiving a worsening of their vocal situation.
In any case, the perceived improvement is present for both the in-clinic and the telerehabilitation group, confirming the non-inferiority of the telerehabilitation treatment compared to the in-clinic one, also in terms of perceived rating scales.

Chapter 4

Conclusions

The analysis conducted in this thesis is based on voice recordings of twenty Multiple Sclerosis (MS) subjects which underwent the Lee Silverman Voice Treatment LOUD (LSVT-LOUD) therapy to treat the symptom of hypophonia. In particular, ten of these patients underwent a LSVT-LOUD treatment conventionally administred in-clinic, while the other group of ten patients underwent the same therapy but innovatively administred in telerehabilitation (Tele-LSVT-LOUD).
The speech material analyzed consists of three repetitions of a sustained vowel /a/, a monologue and a reading recording. These are acquired with a vocal recorder, equipped with an in-air microphone, and with a contact microphone (Vocal Holter, VH), before the therapy (T0), after the therapy (T1) and after three months from the end of the therapy (T2). T2 recordings are not used in this study, since they are not available for all the subjects.
The aim of the analysis is to assess the non-inferiority of the telerehabilitation treatment, compared to the in-clinic one, in terms of various vocal parameters such as Jitter, Shimmer, HNR, CPPS and SPL, and in terms of vocal indexes, that are the Acoustic Voice Quality Index (AVQI) and the Warning Score (WS). This is performed employing different software applications commonly used in the vocal analysis, that are Praat and VOXplot, and also using Matlab.
Therefore, the first part of the analysis deals with the comparison between all the software applications, conducted in terms of AVQImean, computed between all the patients. The AVQI is extracted from a recording made of the concatenation of three seconds of a sustained vowel /a/ and a reading passage. Its computation is conducted considering all the three sustained vowels available for each subject and it involves the parameters CPPS, HNR, Shimmer Local, Shimmer Local in dB, Slope and Tilt.
The results, in terms of AVQImean, highlight a similarity between Praat and VOXplot, confirming the reliability of VOXplot and allowing the use of a practical

and easy-to-handle software application, which provides various parameters and vocal indexes in a few moments and effortlessly.
Instead, from the comparison between Praat and Matlab in terms of AVQImean, it is evident how Matlab and the developed codes, specifically created to analyze the concatenation of a sustained vowel and a reading passage, are also effective in the evaluation of this vocal index.
The comparison between Praat and Matlab is also conducted in terms of SPLmean, computed between all the patients. This confirms the validity of both software applications in the extraction of the SPL. However, Matlab allows to conduct a statistical analysis, returnig the vector of RMS value, while Praat does not.
At this point, the study has the purpose to evaluate the reliability of the contact microphone compared to the in-air one, in terms of WS. The WS is extracted from a sustained vowel recording and it is computed considering the parameters Local Jitter, Local Shimmer, CPPSmedian and CPPSstd. From the results, it is evident that the contact microphone returns values of WS usually lower than those returned by the microphone in air, both at time T0 and T1. This is caused by the fact that the VH returns Local Shimmer values lower than those returned by the microphone in air, often below the threshold that identifies a patological voice. This could be due to the fact that the microphone in air acquires also the environmental noice, being placed at a certain distance from the mouth, which causes random fluctuations in amplitude and, consequently, the Local Shimmer values become higher. For this reason, the contact microphone is considered more precise in extracting the Local Shimmer and so, it is believed to return more reliable values of WS.
Finally, the analisys aims at evaluating the effectiveness of the LSVT-LOUD treatment, as well as the non-inferiority of the telerehabilitation theraphy compared to the in-clinic one, in terms of ΔAVQI and Δ values of some vocal parameters of interest, extracted from the monologue and the sustained vowel recording. For all the software applications, it is highlighted that both the in-clinic and the telerehabilitation therapy produce a vocal improvement, returning values of avereged ΔAVQI and Δ values above zero. Addittionally, the improvements produced following the two therapies are comparable, returning similar values of avereged ΔAVQI and Δ values, as well as overlapping bands of uncertainty. The improvement is evaluated also in terms of WS, that positively decreases for some subjects after the therapy.
All these analysis confirms the reliability of the telerehabilitation therapy (Tele-LSVT-LOUD) compared to the in-clinic one (LSVT-LOUD) and, therefore, the non-inferiority of the telerehabilitation treatment.

4.1 Future works

Even if this work can be considered complete, having demonstrated the non-inferiority of the telerehabiliatation treatment, there are still some aspects that could be treated or further investigated in the future.
Fist of all, since T2 recording are not used in this study due to lack of recordings, it could be necessary to analyze them in the future. This would allow to evaluate the long-term effects of the therapy, understanding if the improvements after the therapy are mantained over time or not. Infact, the risk could be that the improvements are highlighted only immediately after the therapy, and decrease in the follow-up, of three months in this case, leading the patient to return to the initial situation of the disease.
Another purpose could be to evaluate the long term recordings, that are useful for monitoring of the patient's voice during the day. Infact, multiple sclerosis patients often have problems to the voice not only in the execution of specific tasks, but above all in everyday speech, experiencing fatigue. Therefore, the evaluation of the long-term recordings could assess the effectiveness of the therapy also in the daily life of the patient.
Addittionally, the improvement after the therapy could be evaluated with the use of perceptual rating scales draw up by a speech therapist, such as the GIRBAS scale. This can allow to assess the progress of the subjects trough a score assigned by the speech therapist, on the basis of the caracteristics of the voice that are generic grade of dysphonia, instability, roughness, breathiness, asthenia and strain.
Finally, the implementation of the vocal indexes (AVQI and WS) could be integrated in the Vocal Holter device, in order to make it a diagnostic tool, also in terms of these indexes. These computations would be additional functions to those currently available, which only involve the extraction of vocal parameters of interest. This can open up the possibility of immediately and more easily interpreting the patient's clinical situation through the use of vocal indexes.

Bibliography

[1] Zhaoyan Zhang. «Mechanics of human voice production and control». In: *Department of Head and Neck Surgery, University of California, Los Angeles* () (cit. on p. 2).

[2] Mary J. Hawkshaw Robert T. Sataloff Yolanda D. Heman-Ackah. *Clinical Anatomy and Physiology of the Voice*. 1721 Pine Street, Philadelphia, PA 19103, USA: Drexel University College of Medicine (cit. on p. 2).

[3] Jeffrey M. Gelfand. «Multiple sclerosis: diagnosis, differential diagnosis, and clinical presentation». In: *Handbook of Clinical Neurology: Multiple Sclerosis and Related Disorders*. Ed. by D. S. Goodin. Vol. 122. 3rd series. Amsterdam, Netherlands: Elsevier B.V., 2014. Chap. 12, pp. 269–290 (cit. on p. 3).

[4] Nazem Ghasemi, Shahnaz Razavi, and Elham Nikzad. «Multiple Sclerosis: Pathogenesis, Symptoms, Diagnoses and Cell-Based Therapy». In: *Cell Journal* 19.1 (2017), pp. 1–10 (cit. on p. 3).

[5] Valeria Crispiatico, Cinzia Baldanzi, Arianna Napoletano, Laura Tomasoni, Francesca Tedeschi, Elisabetta Groppo, Marco Rovaris, Chiara Vitali, and Davide Cattane. «Effects of voice rehabilitation in people with MS: A double-blinded long-term randomized controlled trial». In: *Multiple Sclerosis Journal* (2021), pp. 1–10 (cit. on p. 3).

[6] C. Vitali et al. «Telerehabilitation for Lee Silverman Voice Treatment (Tele-LSVT)-Loud on voice intensity and voice use in daily living in people with multiple sclerosis: A protocol for a feasibility and pilot randomized controlled study». In: *Digital Health* 9 (2023) (cit. on pp. 3, 4, 7).

[7] Valeria Crispiatico, Cinzia Baldanzi, Arianna Napoletano, Laura Tomasoni, Francesca Tedeschi, Elisabetta Groppo, Marco Rovaris, Chiara Vitali, and Davide Cattaneo. «Effects of voice rehabilitation in people with MS: A double-blinded long-term randomized controlled trial». In: *Multiple Sclerosis Journal* 27.10 (2021), pp. 1–10 (cit. on p. 4).

[8] *Vocal Holter Med Instruction Manual*. Version 2.1.0. 2018, pp. 1–24 (cit. on pp. 10–13).

[9] Paul Boersma and David Weenink. *Praat: doing phonetics by computer [Computer program]*. Version 6.2.06. Retrieved 23 January 2022 from `https://www.praat.org`. 1992–2022 (cit. on p. 14).

[10] Paul Boersma and Vincent Van Heuven. «Speak and unSpeak with PRAAT». In: *Glot Int* 5 (Jan. 2001), pp. 341–347 (cit. on pp. 14, 16).

[11] *VoxPlot*. `https://voxplot.lingphon.com/en/`. Accessed: 20-August-2024 (cit. on pp. 16, 17, 33, 34).

[12] Paul Boersma. «Accurate Short-Term Analysis Of The Fundamental Frequency And The Harmonics-To-Noise Ratio Of A Sampled Sound». In: *Proceedings of the Institute of Phonetic Sciences* 17 (Jan. 2000) (cit. on pp. 20, 22).

[13] *Multi-Dimensional Voice Program (MDVP) Model 5105*. 2005 (cit. on pp. 19, 20, 23).

[14] Jan G. Svec and Svante Granqvist. «Tutorial and Guidelines on Measurement of Sound Pressure Level in Voice and Speech». In: *Journal of Speech, Language, and Hearing Research* 61 (2018), pp. 441–461 (cit. on p. 21).

[15] Soren Y. Lowell, Raymond H. Colton, Richard T. Kelley, and Youngmee C. Hahn. «Spectral- and Cepstral-Based Measures During Continuous Speech: Capacity to Distinguish Dysphonia and Consistency Within a Speaker». In: *Journal of Voice* 25.5 (2011), e223–e232. ISSN: 0892-1997 (cit. on p. 23).

[16] Antonella Castellana, Alessio Carullo, Simone Corbellini, and Arianna Astolfi. «Discriminating Pathological Voice From Healthy Voice Using Cepstral Peak Prominence Smoothed Distribution in Sustained Vowel». In: *IEEE Transactions on Instrumentation and Measurement* 67.3 (2018), pp. 646–654 (cit. on p. 24).

[17] Rubén Fraile and Juan Ignacio Godino-Llorente. «Cepstral peak prominence: A comprehensive analysis». In: *Biomedical Signal Processing and Control* 14 (2014), pp. 42–54. ISSN: 1746-8094 (cit. on p. 24).

[18] J. Hillenbrand and R. A. Houde. «Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech». In: *Journal of Speech and Hearing Research* 39.2 (1996), pp. 311–321 (cit. on pp. 24, 25).

[19] Youri Maryn, Marc De Bodt, and Nelson Roy. «The Acoustic Voice Quality Index: Toward improved treatment outcomes assessment in voice disorders». In: *Journal of Communication Disorders* 43.3 (2010), pp. 161–174 (cit. on p. 26).

[20] Marco Fantini, Andrea Ricci Maccarini, Arianna Firino, Michela Gallia, Vittoria Carlino, Christel Gorris, Massimo Spadola Bisetti, Erika Crosetti, and Giovanni Succo. «Validation of the Acoustic Voice Quality Index (AVQI) Version 03.01 in Italian». In: *Journal of Voice* 37.4 (2023), 631.e1–631.e6 (cit. on p. 26).

[21] Björn Barsties and Youri Maryn. «External Validation of the Acoustic Voice Quality Index Version 03.01 With Extended Representativity». In: *The Annals of Otology, Rhinology, and Laryngology* 125.7 (2016), pp. 571–583 (cit. on pp. 27, 35, 42).

[22] Pasquale Bottalico, Ivano Ipsaro Passione, Arianna Astolfi, Alessio Carullo, and Eric J. Hunter. «Accuracy of the quantities measured by four vocal dosimeters and its uncertainty». In: *The Journal of the Acoustical Society of America* 143.3 (2018), p. 1591. DOI: `10.1121/1.5027816`. URL: `https://doi.org/10.1121/1.5027816` (cit. on p. 27).

[23] Arianna Astolfi, Alessio Carullo, Giuseppina Emma Puglisi, Virginia Isabella Fissore, and Jana Clerici. «An Experimental Approach for the Evaluation of Vocal Behavior of University Professors». In: *Proceedings of the Conference [Nome della Conferenza se disponibile]*. Politecnico di Torino, Italy. Torino, Italy, 2023 (cit. on p. 31).

[24] A. Ricci Maccarini, M. Limarzi, F. Pieri, M. Stacchini, E. Lucchini, and M. Magnani. *La refertazione e l'interpretazione dei tracciati e dei questionari nello studio della disfonia*. Pisa, Italy: Pacini Editore, 2008 (cit. on p. 40).