POLITECNICO DI TORINO

Master's Degree in Biomedical Engineering, Biomedical Instrumentation

Master's Degree Thesis

Vocal Analysis in Multiple Sclerosis Patients Treated with Speech Therapies

A comparison between in-clinic and remote rehabilitation through vocal indexes provided by different software

Supervisors

Prof. Alessio CARULLO

Prof. Alberto VALLAN

Candidate

Alice GIACCHINO

OCTOBER 2024

Abstract

Multiple sclerosis (MS) is a chronic neurological disease that affects the central nervous system, leading to a wide variety of motor, sensory and cognitive impairments. Among these symptoms, speech and voice disorders, such as reduced voice intensity or hypophonia, can significantly affect patients' communication and overall quality of life. Speech therapy can help reduce the grade of disability: one of the most effective methods is the Lee Silverman Voice Treatment (LSVT-Loud), a therapy focused on improving voice intensity. However, it requires frequent clinic visits, which may limit accessibility for some patients. This study conducts an acoustic analysis aiming to evaluate the effectiveness of LSVT-Loud but also showing the potential benefits of tele-rehabilitation (tele-LSVT-Loud) as an alternative approach. Additionally, it compares results obtained from different software commonly used in vocal analysis (Praat, VOXPlot, MATLAB), trying to assess any discrepancies in the parameters extraction. The study was performed on vocal data provided by the MS Rehabilitation Don Carlo Gnocchi Foundation in Milan. The analysis has been conducted on speech material of 20 MS patients, divided in two groups: one group of 10 patients was treated with standard in-clinic LSVT-Loud therapy, while the other 10 MS patients were selected for the Tele LSVT-Loud program, by accessing a telerehabilitation platform from home. Vocal signals were simultaneously acquired using a vocal recorder equipped with an in-air microphone and the Vocal Holter device that embeds a contact microphone. The dataset includes a set of voice recordings provided for each patient before (T0), after the therapy (T1) and during the follow up (T2). Each set includes three repetitions of the sustained vowel /a/, one minute of free speech and a reading of "Notturno", an Italian phonetically balanced text. Two indexes were studied to evaluate the effectiveness of both treatments from T0 to T1: the Acoustic Voice Quality Index (AVQI) and the Warning Score (WS). The AVQI ranges from 0 to 10, with lower values indicating healthier voices. This index includes parameters like jitter, shimmer, Cepstral Peak Prominence Smoothed (CPPS), Harmonic-to-Noise Ratio (HNR), Spectral Slope and Tilt, all extracted by a concatenation of 3 seconds of sustained vowel /a/ and a segment of reading. Using the Praat software, the mean AVQI showed a general improvement across all patients; moreover, the comparison between different software highlighted the

similarity between Praat and VOXPlot, indeed the mean AVQI calculated with both the software differ of 0.14, resulted negligible. A specific MATLAB script was implemented to evaluate the index, focusing on spectral tilt and slope, which differ from Praat respectively of 3.37% and 0.09%. The WS index, calculated on the average of three repetitions of the sustained vowel, also confirmed the effectiveness of both LSVT therapies, as it generally decreased from T0 to T1, indicating an improvement in vocal health. Further acoustic parameters such as HNR, CPPS, f0, jitter, shimmer and Sound Pressure Level (SPL) were analyzed to assess the treatment efficacy, with comparable outcomes between in-clinic and tele-rehabilitation. Finally, parameters extracted from the Vocal Holter were compared to the other software to validate its performance and identify aspects that need to be improved in future configurations.

Acknowledgements

Ringrazio il professor Carullo per la completa disponibilità ed i suoi preziosi consigli durante questi mesi, Giulia Fusari e Chiara Vitali per la gentilezza e l'aiuto durante il nostro lavoro e Davide, per essere sempre stato presente.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Multiple sclerosis (MS) is a progressive neurological disorder that affects the central nervous system, leading to a variety of symptoms that can vary depending on the individual. This chapter provides an overview of MS, examining its pathophysiology, progression, and impact on daily functioning, with a specific focus on speech and voice impairments.
The physiological mechanisms of voice production are then analyzed to illustrate how MS affects the phonatory system.
Finally, a rehabilitation technique, called "LSVT-LOUD", is presented, along with tele-rehabilitation as a potential alternative approach.
This is followed by a presentation of Don Gnocchi Foundation in Milan, which provided us patient data for the analysis.

1.1 Multiple Sclerosis

MS organizations estimate that 2.9 million people in the world have multiple sclerosis [1].
Multiple sclerosis is one of the most common causes of neurological disability in young adults globally: it is a chronic, autoimmune disease that affects the central nervous system characterized by inflammation, demyelination and neurodegeneration, resulting in the impairment of nerve fibre conductivity. It's a complex and unpredictable disorder, for which, nowadays, there is no definitive cure, but there are treatments that can reduce the number and severity of relapses and try to delay

or slow of the long-term progression of the disease, and rehabilitation activities that patients can follow.

The symptoms of MS vary widely from person to person, depending also on the severity of nerve fiber damage and can include fatigue, numbness, vision problems, walking difficulties, making it hard to predict.
Its natural course can be very different, we can mainly distinguish between four types of MS [2]:

- Relapsing-remitting MS (RRMS): characterized by clearly defined attacks of new or increasing neurologic symptoms; these relapses or exacerbations are followed by periods of partial or complete recovery, called remissions. This is the most common type of MS;

- Primary progressive MS (PPMS): defined by worsening neurologic function from the onset of symptoms, without relapses or remissions. Approximately 15% of people with MS are diagnosed with PPMS;

- Secondary progressive MS (SPMS): an initial relapsing-remitting course is followed by an accumulation of permanent disability over time.

- Progressive-relapsing MS (PRMS): the rarest form of MS, including a steady worsening of symptoms from the beginning, with acute relapses

Clinically, motor abilities of a subject affected by MS are identified by a scale (EDSS: Expanded Disability Status Scale), that allows to measure the disability status of the MS patient and monitor its changes over time. The EDDS scale was developed in 1983 by the neurologist John Kyrtzke, and ranges from 0 (that corresponds to a normal neurological exam) to 10 (death due to MS) in 0.5 unit increments that represent higher level of disability. The total score, based on a neurological examination by a clinician, includes partial values of eight functional systems (pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, cerebral or mental and other).

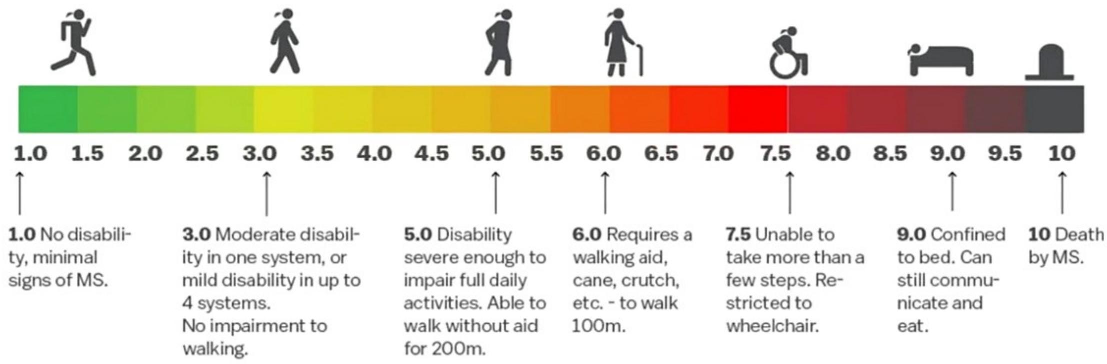

Figure 1.1: The EDSS scale

However, in addition to the motor symptoms caused by MS, cognitive and voice-related symptoms are also very significant: voice and speech are affected in 62% of MS patients [3]; especially patients with greater motor problems tend to exhibit more frequent speech and swallowing difficulties [4].

Phonation is not just a simple exhalation that produces vibrations, but it is a complex physiological process that requires the coordinated involvement of multiple organ systems, in particular the nervous system [5]. Communication disturbances affect the self-confidence of MS patients and their relationships, causing a psychological, social and emotional impact on them.

The main symptoms include dysarthria, dysphonia and hypophonia: the speech can sounds scanning, including long pauses between words, slurred, because of the lack of coordination of the muscles, nasal, harsh, hoarse or breathy. MS patients often can't adapt the volume of the voice according to the environment, and they find difficulties to raise their voce appropriately in noisy settings [6].
Speech therapy, an appropriate rehabilitation and assistive device can help to reduce the impact of this disability; speech-language pathologist can evaluate, diagnose and treat voice disorders, prescribing exercises to enhance functions of the muscles involved in breathing and speaking, teaching techniques to improve speech clarity.

1.2 Voice Signal

The human voice production is complex process that requires four main components:

- a **power source**: lungs, diaphragm, trachea, ribs, chest muscle that are included in the air pressure system;
- **phonation**: the generation of sound by vocal fold vibration;
- **resonance**: the amplification and modulation of the sound by vibration in the structures distal to the larynx;
- **articulation**: complex interaction of lips, teeth, tongue, palate, pharynx to shape and modify the sound.

Vocal folds are two bands of tissue that are positioned opposite each other in the larynx, located between the base of the tongue and the top of the passageway to the lungs, known as the trachea. Each vocal fold is about 11-15 mm long in adult women and 17-21 mm long in men and stretches across the larynx along the anterior-posterior direction [7]. The opening between the vocal folds is called the glottis, which opens during breathing (when the vocal folds are opened and relaxed, allowing air to flow) and narrows during voice production.

The phonation process starts from the adduction of the vocal folds, reducing or closing the glottis (“closed phase”); the contraction of lungs initiates airflow and establishes pressure buildup below the glottis. When the subglottic pressure reaches a certain threshold (to exceed muscular opposition), the glottic slit is forced open (“opening phase”). After air passes through, the subglottic pressure decreases, and the elasticity of the vocal folds, along with the Bernoulli effect (which creates a low-pressure zone as the air flows rapidly through the narrow glottis), pulls the folds back together, initiating the closing phase. This cycle repeats rapidly and produces vibrations, generating a series of air pulses that form sound waves, then shaped into speech by the resonating structures of the vocal tract [8].

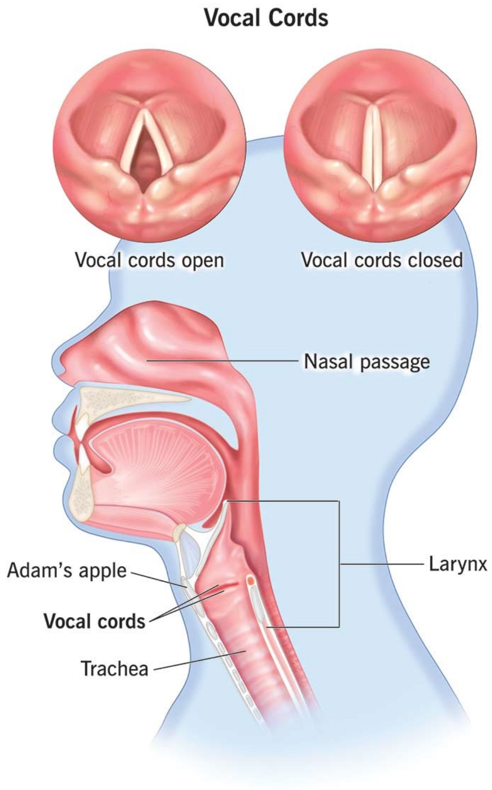

Figure 1.2: Anatomy of phonation cycle, divided into a opening phase and a closing phase.

Voice signal

As described in the previous paragraph, the phonation is the generation of an acoustic signal from a source, the vocal folds. Under normal conditions, the glottal signal, prior to being filtered by the vocal tract and becoming the vocal signal, is a quasi-periodic complex signal, that can be described as sum of sinusoidal signals, called components; each of them is characterized by its own frequency, amplitude and phase. If the signal is periodic with a fundamental period T0, its components ("harmonics") have frequencies that are integer multiples of the fundamental frequency ($f0 = 1/T0$). The fundamental frequency is represented by the frequency of opening and closing of glottis.

The term "*quasi-periodic*" refers to a signal that is sufficiently regular for a human listener to perceive a harmonic sound and identify a fundamental frequency (i.e., vocal pitch) but is not periodic in a strict mathematical sense, i.e. the successive periods are not exactly identical, but differ in frequency and amplitude [9]. The variations, known as short-term perturbations, can occur from one period to the next, or as long-term perturbations, which manifest between the start and the extinction of the sound emission. When these fluctuations exceed certain thresholds, they can characterize a pathological voice, as it will be described in the following chapters.

1.3 Rehabilitation Technique

1.3.1 Lee Silverman Voice Treatment (Tele-LSVT)-Loud

Lee Silverman Voice Treatment (LSVT)-Loud is a highly effective intensive method to treat voice disorders, developed for Parkinson's patients and other neurological conditions, included multiple sclerosis; it aimed at improving respiratory and phonatory effort, vocal fold adduction and laryngeal muscles strength [10]. It helps patients to "recalibrate" their perceptions, enabling them to adjust the volume of the voice according to the external noise levels.

LSTV LOUD traditionally consists in sixteen one-hour therapy sessions across four consecutive weeks and includes daily exercises (from individual words to conversational speech). [11]
In detail, the method consists of high-effort speech exercises, customized according to the patient's functional abilities:

- Synchronous LSVT-Loud sessions:
  - "daily tasks": consist of 30 minutes of prolonged vowel phonation with increasing volume and changing pitch and repetition of 10 functional phrases;
  - "hierarchical exercises": include 30 minutes of reading and conversation exercises with increasing difficulties in duration and complexity of the tasks.
- Autonomous session:
  - 10-minute exercises in addition to the synchronous session four times a week;
  - 15-minute exercises twice a day the remaining 3 days of the week.

It is an efficacious speech treatment for hypophonia, but it typically requires subjects' repeated and periodically attendance at the clinic, which can present significant challenges for individuals with multiple sclerosis: regular travel can be costly and physically demanding, particularly for those with sever motor symptoms, limited mobility or chronic fatigue and non-autonomous patients have to be accompanied by family members or caregiver, making the process more demanding and time-consuming [3]. Moreover, the stress and anxiety associated with in-person / face-to-face sessions and the hospital setting can have a negative impact on the outcomes of the rehabilitation. Furthermore, even though many patients are still workers and independent, integrating these appointments into their daily routine and maintaining regularity is challenging.
All these aspects can make it difficult for patients to consistently and regularly adhere to their rehabilitation program, contributing significantly to treatment discontinuity.

Telerehabilitation can be an alternative method to guarantee easier and continued access to healthcare services, minimising the barriers of distance, time and costs and allowing care to be provided directly at patient's home [12]; furthermore, it may offer personalized, patient-centered care, enhance adherence to therapy protocols and provide an opportunity for real-time monitoring.
Especially with the advent of COVID-19 and subsequent restrictions and containment measures, including social distancing, healthcare systems had to adapt to ensure ongoing patient care: telerehabilitation became essential to avoid the suspension of non-urgent services.

1.4 Don Gnocchi Foundation

This study was conducted in collaboration with the speech therapy and rehabilitation department of Don Gnocchi Foundation in Milan, which provided us with patient data, analysing the speech performance of about 20 patients that suffer from MS.
Don Carlo Gnocchi Foundation is a non-profit, private foundation and one of the biggest in the Rehabilitation, Healthcare, Assistive Technology and Elderly Care in Italy.
Founded in 1952 by Don Carlo Gnocchi to provide care and rehabilitation for wounded children victims of the War, the Foundation has gradually expanded the scope of its operations over the years. In the past half-century, it has mainly dealt with disabled

children which suffer of acquired or congenital pathologies, and patients of all ages who require neurological, orthopaedic, cardiac and respiratory care.
Recognized as an Institution of Hospitalization and Scientific Care (It. IRCCS) in 1991 and as Non-Governmental Organization (NGO) in 2001, today it conducts its activities by relying on the Italian Nation Health Service in twenty-eight residential facilities and thirty local clinics organized in territorial areas [13].

Chapter 2

Materials and methods

2.1 Parameters

Speech alterations can be assessed through the analysis of several parameters extracted from vocal signals with the aim to provide an objective characterization of the voice and identifying specific patterns. The parameters used in the analysis presented in this study are described in the following sections.

2.1.1 Acoustic Parameters

For the definition of the parameters extracted, reference was made to the software instruction manual Model Multi-Dimensional Voice Program (MDVP) Model 5105.

Average Pitch Period - T_0 [ms]

Pitch period (T_0) is defined as the interval between any two successive prominent peaks. For voiced speech, this is the duration of one complete cycle of the vocal fold vibrations that generate the pitch. Average value of all extracted using the pitch period values. Voice break areas are excluded. T_0 is computed from the extracted period-to-period pitch data as:

$$T_O = \frac{1}{N} \sum_{i=1}^{N} T_O^{(i)} \tag{2.1}$$

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

Average Fundamental Frequency - f_0 [Hz]

Analyzing a stationary time signal *x(t)*, the autocorrelation $r_x(t)$ measures the similarity between the signal and a time-shifted version of itself (τ is the lag):

$$r_x(\tau) \equiv \int x(t)x(t+\tau)\,dt \tag{2.2}$$

If the signal is periodic, the lag at which the autocorrelation peaks (exluding the global maxima in zero) is the period T_0; the fundamental frequency of a signal is the inverse of this period.

The average fundamental frequency expressed in Hz is the average value of all extracted period-to-period fundamental frequency values. The voice breaks are ignored.

$$f_0 = \frac{1}{N}\sum_{i=1}^{N} f_0^{(i)} \tag{2.3}$$

where $f_0^{(i)} = \frac{1}{T_0^{(i)}}$ is the period-to-period fundamental frequency and $N =$ is the number of extracted pitch periods.

The fundamental frequency (f_0) is defined as the number of times a sound wave produced by the vocal cords repeats during a given time period; it is determined by the rate of modulation of the vocal folds during voiced speech and it generally decreases in pathological voices. The common frequency ranges are [75-300] Hz for males and [100-400] Hz for females.

Harmonic to Noise Ratio - HNR

The Harmonic-to-Noise Ratio (HNR) measures the relationship between harmonic (periodic) and noise (aperiodic) components of a speech signal, expressed in decibels (dB). It reflects the energy generated by the vocal fold vibrations compared to the noise produced at the glottis, after being shaped by the vocal tract. As described by Boersma [14], the autocorrelation-method is also necessary to calculate the HNR. If the periodic (H(t)) and the noise (N(t)) parts are uncorrelated, the autocorrelation of the total signal equals the sum of the autocorrelations of its parts. Specifically, at zero lag the autocorrelation function is equal to:

$$r_x(0) = r_H(0) + r_N(0) \tag{2.4}$$

If the noise $N(t)$ is white (i.e., uncorrelated with itself), the autocorrelation function $r_x(\tau)$ will exhibit a local maximum at lag $\tau_{\max} = T_0$ with a peak value of:

$$r_x(T_0) = r_H(T_0) \tag{2.5}$$

Since the autocorrelation function at zero lag represents the total power of the signal, the normalized autocorrelation at $\tau_{\max}$ indicates the relative power of the periodic (or harmonic) component of the signal, while the complement represents the relative power of the noise component.

$$r'_x(\tau_{\max}) = \frac{r_H(0)}{r_x(0)} \tag{2.6}$$

$$1 - r'_x(\tau_{\max}) = \frac{r_N(0)}{r_x(0)} \tag{2.7}$$

The HNR can therefore be defined as follows:

$$HNR\,(\mathrm{dB}) = 10 \cdot \log\left(\frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})}\right) \tag{2.8}$$

Cepstral Peak Prominence Smoothed - CPPS [dB]

According to the definition given by Hillenbrand and Houde [15], the cepstral domain is a log power spectrum of a log power spectrum: the first spectrum shows the frequency distribution of the signal energy, while the second indicates how periodic the harmonic components in the spectrum are.
Mathematically, the Cesptrum is defined as:

$$\mathcal{C}_p(\tau) = \left|\mathcal{F}\left\{\log\left(|\mathcal{F}\{x(t)\}|^2\right)\right\}\right|^2 \tag{2.9}$$

where $x(t)$ is the voice signal, $\mathcal{F}$ is the Fourier transformation, $|\mathcal{F}\{x(t)\}|^2$ is the signal power spectrum and tau is the quefrency.
For a periodic signal, the first power spectrum will show energy at harmonically related

frequencies; performing a second Fourier transform on it, the cepstrum obtained highlights a strong component corresponding to the regularity of the harmonic peaks. The "quefrency" (1/frequency) at the cepstral peak corresponds to the fundamental period of the signal and the amplitude of this peak represents both the level of harmonic organization and the overall signal amplitude. This peak is located between the minimum and the maximum expected fundamental periods [16].
Two cepstral parameters have been defined: the Cepstral Peak Prominence (CPP) and its smoothed version, the Cepstral Peak Prominence Smoothed (CPPS). These measures can be used to analyze dysphonic voices during both connected speech tasks and sustained phonation.
The CPP is the measure of the amplitude of the cepstral peak corresponding to the fundamental period, normalized of overall signal amplitude through a linear regression line estimated relating quefrency to cepstral-magnitude (computed between 1 ms and the maximum quefrency): [17] in other words, it is the difference between the cepstral peak and the corresponding value on the regression line.
The concept of the CPP measure is that a signal with a high degree of periodicity will exhibit a defined harmonic structure, leading to a more prominent cepstral peak and an higher value of CPP compared to a less periodic signal (as a pathological, breathy voice) [15].
The CPPS considers two smoothing steps before calculating the cepstral peak prominence, averaging the cepstral magnitude across time and quefrencies before extracting the cepstral peak.
First, the cepstra are averaged across time: each unsmoothed cepstrum is replaced by the average of some cepstral frames both preceding and following the current frame. Similarly, in the second step the smoothing is performed in the quefrency domain, using an averaging rectangular window. On the obtained cepstrum, is computed the regression line and is searched the peak.
The CPPS mean is the average of the cepstral peak prominences of the individual frames.

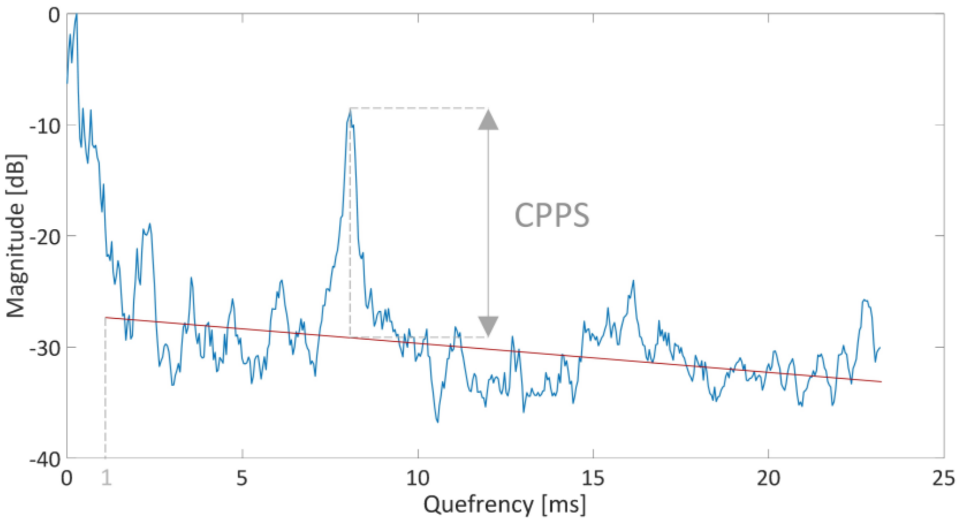

Figure 2.1: CPPS

Sound Pressure Level - SPL

The air pressure variations due to the sound propagation are commonly referred to as "sound pressure". The sound pressure can be expressed in the unit of Pascal, but it is more commonly expressed using an SPL in decibels [18]. The Sound Pressure Level (SPL) quantifies the intensity and the loudness of a sound, estimating the amount of acoustic energy and power radiated out of the mouth.
It is a logarithmic measure of the effective pressure of sound relative to a reference value; the used reference sound pressure in air is 20µPa, which is considered the threshold of human hearing and corresponds to a SPL of 0 dB.
The SPL (in dB) is defined in IEC 61672-1 (IEC, 2013) by the following formula:

$$L_p = 10 \log \left(\frac{p}{p_0} \right)^2 \tag{2.10}$$

where L_p is the SPL (Sound Pressure Level) expressed in dB, p is the sound pressure in Pascal and p_0 is the reference sound pressure (20 µPa).

In order to measure SPL correctly, it is important to use calibrated equipment, ensuring that the levels are related to the reference pressure of 20 µPa. To measure SPL parameter with the Vocal Holter, a preliminary calibration is required: each subject has to repeat the vowel /a/ at progressively increasing volume levels in front of the embedded in-air microphone integrated into the DAP unit, located at 22 cm from the mouth of the patient; the same in-air microphone was previously calibrated against a pressure calibrator, which provides an SPL level of 94 dB via a signal at a frequency of 1 kHz.

Also a calibration of the in-air microphone (located at 30 cm from the subject's mouth) is necessary, in order to create a function to convert RMS values of the sound signal into the amplitude in dB, corresponding to the SPL parameter; this formula has been used both in Matlab and Praat scripts:

$$p = 0.0027 + 6.0474 \cdot \mathrm{RMS} \tag{2.11}$$

where

- p is the sound pressure (Pa) on the microphone plane
- RMS is defined as the square root of the arithmetic mean of the squares of the voice samples:

  $$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x_n^2} \tag{2.12}$$

  where x_n are signal samples. The RMS value is also employed in the pre-processing phase in the Matlab scripts to discriminate the voiced and unvoiced frames.

The SPL values are higher when measured closer to the mouth and decrease with the increasing distance. Therefore, it is essential to report the measurement distance together with the measured levels of voice. A standard distance of 30 cm is recommended, ensuring that the voice signal is sufficiently amplified to ambient noise and minimizing errors caused by subject's movements. [18]
To facilitate comparison with values typically reported in the literature, this SPL measurement is converted to a standard reference distance of 1 meter; to specify the parameter SPL at a different distance, the following expression can be used:

$$SPL_d = SPL_{d_0} + 20\log_{10}\left(\frac{d_0}{d}\right) \tag{2.13}$$

where SPL_d is the SPL referred to the distance d, SPL_{d_0} is the SPL calculated at

a distance d_0 from the mouth, d_0 is the reference distance at which the SPL was initially measured, d indicates the generic distance at which you want to calculate the SPL.
For instance, to estimate the sound pressure level at a distance of 1 m from a sound source, given a measurement taken by the air-microphone at 30 cm, the inverse square law can be employed:

$$\mathrm{SPL}_{1m} = \mathrm{SPL}_{30cm} + 20\log_{10}\left(\frac{30}{100}\right) = \mathrm{SPL}_{30cm} - 10.458 \tag{2.14}$$

2.1.2 Stability Parameters

Jita [µs] : absolute Jitter

It is an absolute measure of the period-to-period variability of the pitch period within the analyzed voice sample, used on voice perturbation. Voice break areas are excluded.

$$Jita[\mu s] = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_0^{(i)} - T_0^{(i+1)}\right| \tag{2.15}$$

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

Jitt [%]: Jitter Percent

Relative estimation of the variation of the period-to-period variability of the pitch.

$$Jitt[\%] = 100 \cdot \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_0^{(i)} - T_0^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \tag{2.16}$$

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

Both Jita and Jitt represent evaluations of the same type of pitch perturbation: period irregularity can be associated with the inability of the vocal cords to sustain a periodic vibration for a defined period, causing an hoarse voice.
While Jita is an absolute measure, depending on the average fundamental frequency

of the voice and so differing significantly between men and women, Jitt is a relative measure that facilitates comparisons between subjects.
High jitter values indicate irregularities in the vocal fold vibration.

ShdB [dB]: Shimmer in dB

It is a measure of the period-to-period variability of the peak-to-peak amplitude within the analyzed voice sample. Voice break areas are excluded.

$$ShdB[dB] = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \log \left(\frac{A^{(i+1)}}{A^{(i)}} \right) \right| \tag{2.17}$$

where $A^(i)$, $i = 1, 2, \ldots, N$ are the peak-to-peak amplitude data and N the number of extracted impulses.

Shimm [%]: Shimmer Percent

Relative evaluation of the very short term (cycle-to-cycle) irregularity of the peak-to-peak amplitude of the analyzed voice sample. Voice break areas are excluded.

$$Shim[\%] = 100 \cdot \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A^{(i)} - A^{(i+1)}|}{\frac{1}{N} \sum_{i=1}^{N} A^{(i)}} \tag{2.18}$$

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

Both Shim and ShdB are relative evaluations of the same type of amplitude perturbation, but they use different measures for the results. Cycle-to-cycle variations in amplitude can be associated with the vocal cords' inability to sustain periodic vibration and with the presence of turbulence noise in the voice signal; the voice can sound hoarse and breathy.
High shimmer values suggest variations in the loudness of the voice.

2.2 Indexes

2.2.1 AVQI

Maryn et al. developed the Acoustic Index Quality Index (AVQI), a multiparameter indicator to obtain a single score for objectively estimating overall voice quality and measuring dysphonia severity to monitor treatment outcomes.

The clinical utility of AVQI has been evaluated by determining its responsiveness to change, that is the ability of a measure to detect change after an intervention or a rehabilitation treatment.

AVQI is the first measure to incorporate samples of continuous speech, in addition to the sustained vowel samples used in other measurement protocols: in fact, it is calculated using the concatenation of 3 seconds of vowel [a:] and the voice segment reported below, extracted from a connected speech (a phonetically balanced text, that incorporates a weighted combination of metrics from the time, frequency, quefrency domains).

"*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.*"

Including both these speaking tasks in voice analysis is essential, because dysphonia symptoms and vocal inconsistencies, such as voice breaks or prosodic modulation, are often more pronounced during conversational speech rather than sustained vowels; furthermore, the two stimulus types can express different types or degrees of vocal dysfunction, so each of them can be decisive in a voice quality evaluation [19].

AVQI is computed using a specific algorithm based on linear regression analysis that combines six multiple acoustic parameters: CPPS, HNR, Shimmer local, Shimmer local dB, general slope of the spectrum and tilt of the regression line through the spectrum.

$$\begin{aligned} \mathrm{AVQI} = [&4.152 - (0.177 * \mathrm{CPPS}) - (0.006 * \mathrm{HNR}) - (0.037 * \mathrm{Shimm}) \\ &+ (0.941 * \mathrm{ShdB}) + (0.01 * \mathrm{Slope}) + (0.093 * \mathrm{Tilt})] * 2.8902 \end{aligned} \tag{2.19}$$

The following list provides a detailed overview of the parameters used in Praat for determine AVQI:

Table 2.1: List of parameters'Praat definitions to calculate AVQI

Time Domain	
HNR	Base-10 logarithm of the ratio between the period energy and the noise energy multiplied by 10
Shimmer Local	Absolute mean difference between the amplitudes of successive periods, divided by the average amplitude
Shimmer Local dB	Base-10 logarithm between the amplitudes of successive periods multiplied by 20
Frequency Domain	
Slope	Difference between the energy in the 0-1 kHz range and the energy in the 1-10 kHz range of the long-term average spectrum
Tilt	Difference between the energy in the 0-1 kHz range and the energy in the 1-10 kHz range of the trendline through the long-term average spectrum
Quefrency Domain	
CPPS	Distance between the first harmonic's peak and the point with equal quefrency on the regression line through the smoothed cepstrum

The AVQI ranges between 0 and 10, with a smaller number indicating healthier voice and an higher one correlates with a more dysphonic voice quality, but each language has its own cut-off value and diagnostic accuracy: for example, the optimal cut-off point to discriminate between normophonic and dysphonic voices for Italian was identified at AVQIv3-IT < 2.35, similar at French (2.33), instead for German is at 1.85, and for Korean 3.15 [20].

In this thesis, in order to calculate AVQI scores, some scripts in Matlab were implemented, and a modified version of AVQI 03.01 Praat script by Youri Maryn, PhD and Paul Corthals PhD was used.

This script enabled two separate analyses: first, it facilitated the comparison between the Voxplot and Praat software. Second, it allowed us to evaluate the effectiveness of the rehabilitation by calculating the AVQI using Praat, assessing the AVQI levels at two distinct time points: at the beginning of the rehabilitation, T0, and at the end of the program, T1, for both in-person and home-based rehabilitation.

2.2.2 Warning Score

The Warning Score is an index that can assess the vocal health status of the subjects, indicating potential vocal diseases. It can be used to flag abnormal results during voice monitoring, suggesting that the vocal signal deviates from expected norms, and further investigation may be needed.
The Warning Score is included in the range (-4 ÷ +4), with particular attention needed for subjects with positive values, because they can be pathologic.
It is calculated only on the repetition of the sustained vowel /a/, including the values of Jitter (%), Shimmer (%), median and standard deviation of CPPS (dB); depending on these parameters, voice recordings are classified as “healthy”, “pathological” or “not reliable”.

Each parameter contributes to the final value of the WS, implementing these rules, provided from [21] :

Table 2.2: Values of contributes to WARNING SCORE

Parameters	**Value**	**Voice identified as**	**CONTRIBUTION TO WS**
Jitter (%)	<0.31	Healthy	-1
	>0.43	Pathological	+1
	0.31 ÷ 0.43	Not Reliable	0
Shimmer (%)	<2.37	Healthy	-1
	>2.55	Pathological	+1
	2.37 ÷ 2.55	Not Reliable	0
CPPS median (dB)	>19.7	Healthy	-1
	<18	Pathological	+1
	18.0 ÷ 19.7	Not Reliable	0
CPPS std (dB)	<0.9	Healthy	-1
	>1.3	Pathological	+1
	0.9 ÷ 1.3	Not Reliable	0

The final warning score is the sum of the single contribution of each parameter.

2.3 Software

In this chapter, the software utilized for voice analysis in this study will be introduced.

Three distinct tools were employed: MATLAB R2024a, a commercial software for numerical computation and data analysis, Praat, a free and open-source software widely used in audio and speech analysis, and VOXPlot, a custom software specifically developed for voice analysis.

The same dataset was analyzed using all three software applications, allowing for a comparison of the similarities and differences observed throughout the process, from preprocessing to parameter extraction. The outcomes were then compared to evaluate the performance and accuracy of each software.

2.3.1 Praat

"*Praat: doing phonetics by computer.*" [22]

It is a free, open-source software package written and maintained by Paul Boersma and David Weenink at the Institute of Phonetic Sciences of the University of Amsterdam, principally used for speech analysis, speech synthesis, speech manipulation; it is compatible with a variety of operating systems, including Unix, Linux, Mac and Microsoft Windows.
When the program starts, two main windows appear:

- the object window: the main workspace where the user can manage and interact with the data items; it displays the list of the objects (created or imported in Praat), analysis results and includes a dynamic menu with commands available for the selected object;
- the picture window: used to create, edit and export graphs and visual representations of the data.

The principal functions are: [23]

- Analysing audio: the sound window displays the waveform, the spectrogram, the pitch contour and the formant contours, enabling also users to measure pitch, intensity, duration, formants, and other acoustic parameters;
- Annotating speech: PRAAT is widely used by linguists to label and segment speech recordings, obtaining transcriptions and annotations;

- Synthesizing speech: it allows users to create simple or more complicated sounds, also performing source-filter and articulatory synthesis (generate sounds from a specification of acoustics parameters or timed muscle contractions);

- Manipulating speech: it is also possible to adjust pitch and duration contours of the signal and apply different filtering techniques.

With Praat's scripts is possible to automate and standardize most of the sound formatting and acoustic analyses.

2.3.2 VoxPlot

VOXPlot is an open-source, multilingual software based on Praat (version 6.0.48) algorithms.
It was developed by Dr. Jorg Mayer (in close cooperation with scientific advisor Prof. Ben Barsties v. Latoszek) for the Acoustic Voice Quality Analysis (AVQI), making it standardized and simple, with an interface that meets the demands of intuitive ease of use for clinicians and researchers [9].

The start screen features two windows, allowing users to load or record two different signals (a continuous speech and a sustained vowel of at least 3sec), insert patient's data and details of the examination, choose the analysis language and perform a vocal analysis on them. VOXplot only processes audio data in WAV format, other audio formats, such as MP3 files, are not suitable; voice samples recorded in VOXPlot are automatically saved in WAV format. When both a continuous speech (CS) sample and a sustained vowel (SV) sample of 3 seconds are available, a full analysis is conducted. If only one sample is present, or if the SV sample's duration is insufficient or exceeds the required length, a warning is displayed. In such cases, the analysis can either be canceled to correct the samples or proceed as an incomplete analysis.

For a complete voice analysis, VOXPlot requires two voice samples:

1. a reading passage (called continuous speech, "cs"), which is selected according to the analysis language.

2. 3 seconds of the sustained vowel /a/ ("sv"): if the *sv* voice sample is shorter than 3 seconds, a warning will indicate that it is unsuitable for standardized AVQI or ABI analysis; nevertheless, it will still be loaded and remains available

for analysis. For *sv* voice samples longer than 3 seconds, they can be easily shortened by adjusting the selection box.

VOXPlot, as the outcome of the vocal analysis, generates a voice profile, which is a concise PDF (or JPG/PNG) sheet that includes all examination data, measured values, a narrowband spectrogram for the sustained vowel and a norm-value circle that highlights deviations in 6 acoustic dimensions using an intuitive traffic light system.

The parameters analysed are calculated for the "sv", for the "cs" and for a combination of both samples (called "mx"), that includes complete the entire sv sample and the phonated parts of the cs sample:

- Spectral slope (dB)
- Spectral tilt (dB)
- Amplitude difference between 1st and 2nd harmonics in the spectrum (H1H2) (dB)
- Smoothed Cepstral Peak Prominence (CPPS) (dB)
- Frequency perturbation: Jitter Local (%) and Jitter ppq5 (%)
- Amplitude perturbation: Shimmer (%) and Shimmer (dB)
- Harmonics-to-Noise Ratio (HNR) (dB)
- HNR Dejonckere & Lebacq (HNR-D) (dB)
- Period Standard Deviation (PSD) (ms)
- Voice breaks
- Glottal-to-Noise Ratio (GNE)
- High-Frequency Noise (HF noise)
- Pitch and its related parameters (Pitch mean, pitch min, pitch max, pitch sd, pitch range)
- Acoustic Voice Quality Index (AVQI)

- Acoustic Breathiness Index (ABI)

AVQI and ABI are two multiparametric acoustic indices that are calculated on the combination of the two voice samples available (continuous speech and sustained vowel).

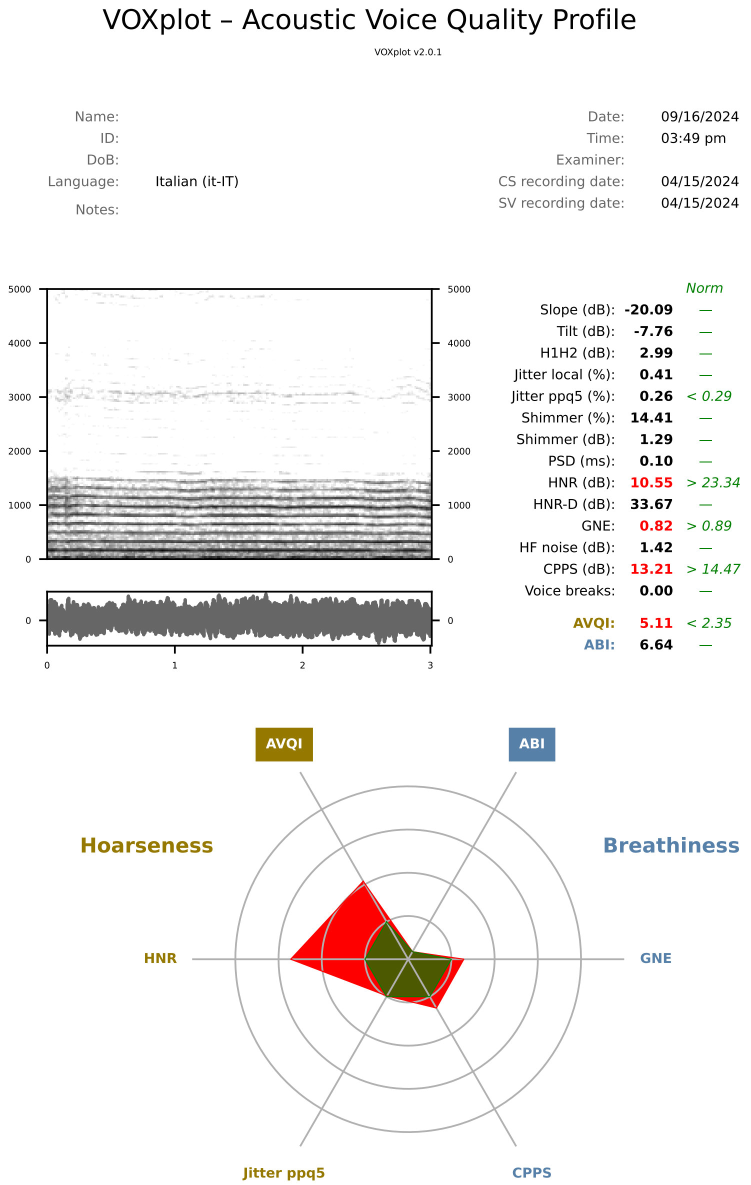

Figure 2.2: Voice profile generated by VOXPlot, as outcome of vocal analysis

2.4 Dataset

The dataset is composed by 20 multiple sclerosis (MS) patients, recruited from the MS Rehabilitation Unit of the IRCSS Don Carlo Gnocchi Foundation ONLUS of Milan (Italy); 10 MS patients were randomly selected to form the control group and follow the standard rehabilitation program (LSVT-Loud in the clinic), while the remaining 10 patients were treated with the telerehabilitation therapy (Tele-LSVT-Loud). For the latter, it is used a digital rehabilitation platform (Maia Platform, https://abmedica.it/prodotti-ab-medica/maia), which patients could access via their personal PC at home, accessing a telepresence system (videoconference).

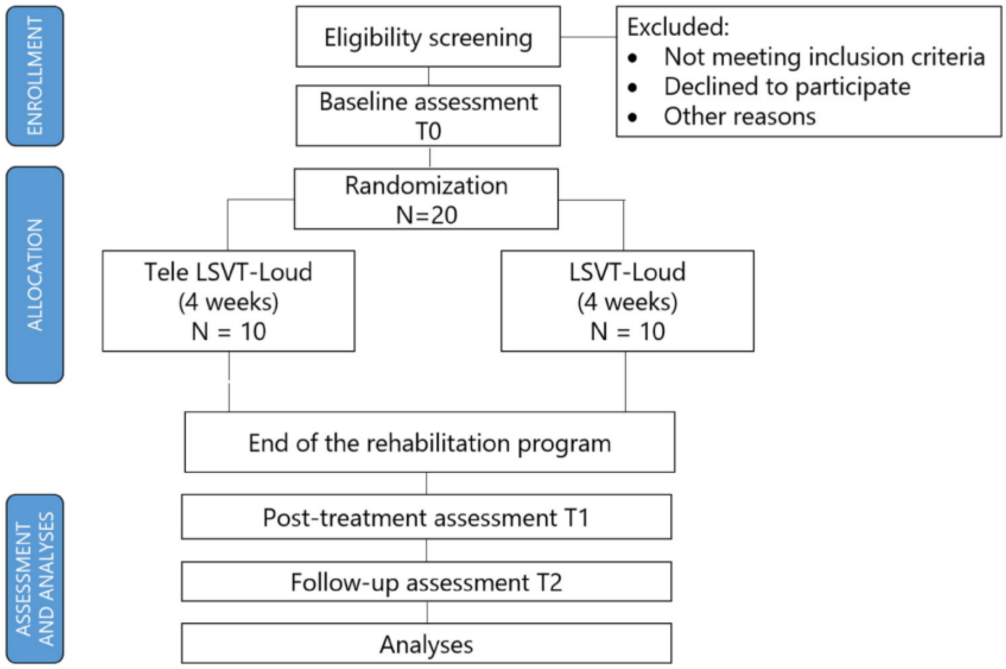

Figure 2.3: The protocol trial

As reported in the related protocol [3], the patients selected for the study have to be included in these inclusion criteria:

1. Age greater than 18;
2. Diagnosis of multiple sclerosis based on McDonald's criteria;
3. Presence of mild-to- severe voice symptoms, as confirmed by two speech-language

therapists;

4. Mini-Mental State Examination score greater than 24[40];

5. Availability of a personal computer and internet connection at home;

6. Stable pharmacological treatment with dopamine agonists and/or steroids in the last 3 to 6 months, if any;

7. Agreement to participate in the study, with the signature of the informed consent form.

The exclusion criteria include:

1. Presence of dysphonia related to other diseases;

2. Presence of comorbidities such as other neurological conditions different from multiple sclerosis;

3. A history of laryngeal cancer, radiotherapy, head or neck trauma, or intubation;

4. Presence of visual or hearing impairments;

5. Presence of major psychiatric comorbidities;

6. Participation in voice rehabilitation sessions using conventional treatment or LSVT-Loud in the last 6 months.

For each patient is provided a set of voice recordings acquired at baseline (T0), after treatment (T1) and during the follow-up (T2, 3 months after T1). At the time of this study, T2 data are not still available, so the analysis focuses only on T0 and T1 recordings.
The speech material includes:

- Three repetitions of the vowel /a/ at a comfortable pitch, level and duration;

- One-minute of free speech, telling something they knew well;

- Reading of “Notturno”, an Italian phonetically balanced text (reported in the appendix).

The vocal signal are recorded using a portable high-quality (24 bit/96 kHz) recorder with an integrated in-air microphone (Roland, R-05) and the device Vocal Holter (VH).

Patient	Gender
Patient 1	F
Patient 3	F
Patient 5	F
Patient 6	M
Patient 8	F
Patient 11	F
Patient 13	F
Patient 15	F
Patient 19	F
Patient 20	M

Table 2.3: LSVT-Loud in clinic

Patient	Gender
Patient 2	F
Patient 4	M
Patient 7	M
Patient 9	F
Patient 10	F
Patient 12	F
Patient 14	F
Patient 16	F
Patient 17	F
Patient 21	F

Table 2.4: Tele-LSVT-Loud

2.5 Data acquisition

Vocal recordings of the involved subjects were simultaneously acquired using a vocal recorder equipped with an in-air microphone and the Vocal Holter device that embeds a contact microphone. The samples are saved in .wav format.

2.5.1 In-air microphone

The portable high-quality recorder used to record the acoustical tasks of this study was the Roland R-05 vocal recorder with an in-air microphone, designed to record a wide range of sounds.

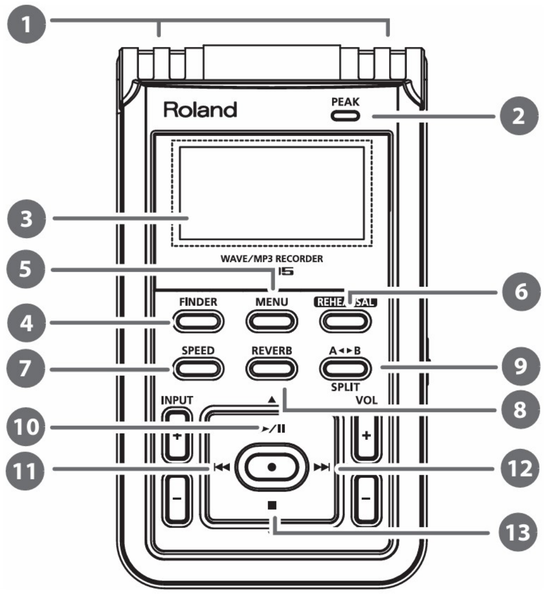

Figure 2.4: Roland R-05 vocal recorder, where:
1) stereo microphone; 2) peak indicator, that indicates if input/output volume is excessive; 3) display; 4) FINDER button, used to view, rename, delete files; 5) MENU button, to modify settings or specifying date and time; 6) REHERSAL button, that automatically sets the appropriate recording level; 7) SPEED button to change the playback speed; 8) REVERB button, to turn reverb on/off; 9) SPLIT button; 10) Start/pause a recording; 11) Rewind/select the previous file; 12) Fast-forward/ select the next file; 13) Stop button

The default sampling rate is 44.1 kHz, and the recordings are saved in .wav-16 bit files. The built-in microphone are stereo: the mic on the right records the R channel, and the mic on the left side records the L channel. To record patient's tasks, the microphone has to be point diectly at the center of the his face so that the recording isn't unbalanced toward left or right channel.

2.5.2 Vocal Holter

Vocal Holter Med (VHM or VH) is a monitoring system based on Voice Care® technology, designed by Polytechnic University of Turin, in collaboration with the S.C. ORL 2 U. of the University of Turin and PR.O.VOICE srl, and it is configured as an electromedical device to aid in the diagnosis of disorders of the vocal apparatus. It's a non-invasive, wearable device, that allows to monitor the quality of the voice during the time (both short and long-term evaluation) and to avoid the effect of the environmental noise on voice signal acquisition, being useful both as primary prevention tool and as diagnostic instrument.

Figure 2.5: The elements of the kit Vocal Holter Med

The kit, shown in figure 2.5, is composed of:

- **DAP: Data Acquisition and Processing unit**, which includes:
  - **Audio microphone**: Used during the calibration of the contact microphone as a reference and during long-term evaluation to estimate background noise.

– **Telescopic antenna**: Used as a spacer during the calibration of the contact microphone.

- **Contact microphone (model hx-505-1-1)**.
- **Power adapter and cable**.

(a) Use of the telescoping antenna during the calibration of the contact microphone

(b) Correct positioning of collar of contact microphone around the patient's neck

Figure 2.6: Vocal Holter Med device worn

Unlike the air-microphone, the VHM is not affected by other acoustic sources, as physical activity, distance and orientation from the mouth during the test: in fact, there is a piezoelectric contact microphone, positioned at the neck at a fixed distance with a collar, that measures the skin vibrations caused by the activity of the vocal cords and a device, that can be worn in a pocket, that allows to acquire, process and store the data. The contact microphone is not negligible to body movements, but a digital low-pass filter (implemented in the VHM firmware) select the frequency components related to voice activity, discarding the ones related to body movements.

VHM is able to measure:

- **Vocal parameters:**
  - Sound Pressure Level (SPL) at approximately 22 cm from the mouth [dB].
  - Fundamental Frequency (F0) [Hz].
  - Voicing time percentage (Dt%).

  - Local Jitter [%].
  - Local Shimmer [%].
  - Cepstral Peak Prominence Smoothed (CPPS) [dB].

- **Environmental parameters:**
  - Background Noise Level (LA50, LA75, LA90, LAeq).
  - Air Temperature (θ) [°C].
  - Air Relative Humidity (h) [%UR].

The Vocal Holter sample the signal at 44.1 kSa using 16-bit resolution: the samples acquired are then grouped into frames of 46ms. The parameters are calculated only on the speech frames; to obtain the values measured during the recording, the user has to connect to the web interface, that includes different sections:

- **Calibration section**: Used to start the procedure for calibrating the two microphones (the audio and the contact microphone).
- **Evaluation section**: Allows the user to choose between a short-term or long-term evaluation.
- **Environmental and system section**: Displays the current values of temperature and relative humidity, as well as the level of battery charge.
- **Download section**: Enables the downloading of parameters and data.

Two types of evaluations can be performed:

Short-time evaluation

The short-time evaluation is used during the sustained vowel /a/; if the subject produces more than one vocalization during the total recording time, only the first one will be processed. At the end of the recording, the DAP unit processes the data and a message window displays the values of the following parameters: fundamental frequency, local jitter, local shimmer, CPPS mean and CPPS standard deviation;

```
Short evaluation started on date: 2024-04-08 09:05:17
Calibration parameters: Mean squared error: 3.55 dB, linearity: 81.2 %, Intensity ratio: 0.8 : 1
Battery charge: 84 (%) Temperature: 26.2 (C) Relative humidity: 63 (%RH)


Fundamental frequency: 109.3 Hz
Jitter: 3.41 %
Shimmer: 7.75 %
CPPS (median): 14.4 dB
CPPS (standard deviation): 4.72 dB
```

Figure 2.7: Example of VH short-time evaluation

Long-time evaluation

This method involved monitoring during the reading of a passage of a text, a free speech or a lecture. The results can be saved in various formats. The "Standard file" format includes detailed information such as the date and time of the recording start, calibration parameters, battery charge, temperature, relative humidity (%RH), and background noise level (BNL) along with its statistical measures (LAF90, LAF75, LAF50, Leq). Additionally, it captures voicing time percentage (Dt%), sound pressure level (SPL) with its statistics (mean, median, standard deviation, 5th and 95th percentiles), and fundamental frequency with corresponding statistics (mean, median, standard deviation, 5th and 95th percentiles). These parameters are updated approximately every 75 seconds. The "Tab-separated-file" format presents the data in a tabular layout with three rows and nine columns. The rows contain the date and time of the evaluation start, calibration details, and environmental parameters. The columns include time (with values updated every 75 seconds), battery charge, temperature, relative humidity, BNL and its four statistics, voicing time percentage, and statistical measures of SPL and fundamental frequency. Lastly, the "Tab-separated-file (46ms)" format starts with the same three rows as the previous format but includes two additional columns for fundamental frequency (f0) in Hz and SPL in dB. These values are acquired and updated every 46.4 milliseconds.

In this dataset the audio weren't recorded through the Vocal Holter, so it wasn't possible to download the voice samples. From the data returned by the DAP unit, as "standard file" (Fig. 2.7) and "tab-separated file", parameters were automatically extracted through a Matlab script.

The aim of this study is to compare the VH parameters with those extracted from the in-air microphone and validate the use of VH as a tool to assess voice quality and the degree of disease progression through indices such as the warning score. This comparison was conducted by evaluating the parameters of the same audio files, but extracted from both microphones: a Δ for each parameter was calculated, defined as the difference between the parameter value measured with the air microphone and the value obtained from the DAP unit of the VH.

2.6 Preprocessing

Firstly, a manual cleaning has been necessary: once the dataset of voice recordings was loaded, empty or unsuitable recordings, which were considered "faulty" or poorly recorded, were removed.
The recordings were provided in two separate .wav stereo files: one containing the three repetitions of the sustained vowel, and another including both the reading task and the monologue. Using Audacity software (version 3.5.1), these recordings were manually trimmed to extract the individual segments and were consistently labeled with the same naming convention for each patient to facilitate the automation of the subsequent scripts (respectively *"sv1", "sv2", "sv3", "let", "mon"*). During this step, initial and final silences were removed, along with irrelevant sections such as instructions for the task and background noise.
The result of this manual cleaning are five different audio files:

- *3 vowel /a/ files*: three sustained vowel recordings, cut to 3 central seconds, where the signal is more stable;
- *reading file*: the central reading part of "Notturno" passage;
- *free speech file*: 60 seconds of free monologue.

The cleaned data are then processed in different software, with different pre-processing and functions to extract the parameters.

The audio recordings were captured using an in-air micropone, which produces two-channel (stereo) recordings, typically used to record sounds with spatial information, including left and right channels. However, for vocal analysis, using only one-channel (mono) is preferred to focus on the voice itself, rather than the spatial information and because the slights differences between the two channels can complicate the analysis or introduce non relevant noise. Extracting only one channel (the first) simplifies and speeds up the process of acoustic analysis. In MATLAB and Praat, the extraction of the single audio channel was implemented at the beginning of the scripts, whereas for VOXPlot, it was done manually before loading the files.

2.7 Feature extraction from different software

2.7.1 Praat

Below is a detailed description of the various steps involved in the script's functionality:

Input

The script runs on two recordings, "sv" (3 seconds of sustained vowel /a/) and "cs" (continuous speech, a segment of reading of "Notturno", selected to be the same for each patient);

Preprocessing

- **Extraction of the first channel** from each stereo sounds;
- **Filtering**: a 0-34 Hz notch filter is applied to the recordings to remove low-frequency noise, artifacts, and any potential direct current (DC) components;
- **Detection, extraction and concatenation of the voiced segments in the recording of continuous speech**: the script first generates a TextGrid (a Praat object used for annotation of segmentation and labelling) to identify silences in the recording, then extracts and combines the non-silent portions into a new sound object (called "onlyLoud"). The silence threshold, set to -25 dB, determines the maximum silence intensity value in dB with respect to the maximum intensity; intervals with an intensity smaller than the threshold for a minimum duration of 46 ms (to align it at Matlab script) are considered as silent intervals. Using a sliding window, it analyses the power of each segment, applying a threshold to exclude those with insufficient power, and calculate the zero-crossing rate. Only segments with a partial power that exceeds the threshold (30% of Global Power) and a zero-crossing rate higher than a certain value (3000) are considered as "voiced" and are concatenated into a single sound file, called "onlyVoice";
- **Concatenation** of a 3-second sustained vowel to the "onlyVoice" sound file.

Parameters extraction

The parameters extraction includes the determination of the six acoustic measures calculated on "onlyVoice" sound:

- **CPPS**: first the PowerCepstrogram of the sound is obtained using the command "*To PowerCepstrogram... 60 0.002 11025 1*": the sound is resampled to twice the value of the Maximum frequency (set to 11025 Hz, to align it with Matlab), then a pre-emphasis is applied from 1 Hz; for each analysis window, determined by the parameter "pitch floor" (the length of the analysis window is determined as three periods of this pitch, 3/60 = 0.05 s), a Gaussian window is applied to obtain the spectrum. Applying a second Fourier transform to the logarithm of the power spectrum, a PowerCepstrum is obtained. Finally, the values of the PowerCepstrum are stored in vertical slice of the PowerCepstrogram, a time-quefrency representation of the sound. "*Get CPPS*" command allows to obtain the value of CPPS from the PowerCepstrogram, as the average of the cepstral peak prominence of the individual frames.

  The smoothing of the PowerCepstrogram is performed using a time averaging window of 0.014s and a quefrency averaging window of 0.001s; the result of this step is a new smoothed PowerCepstrogram, where each cepstral value is the average of the cepstral values within the averaging window positioned symmetrically around the center of each frame, both in time and quefrency domain.

  The peak search is conducted in the range [60 - 330] Hz, with a tolerance of 0.05 and a parabolic interpolation; once the cepstral peak is identified, the algorithm compares it to a baseline, calculated by fitting a straight trend line between 1 ms and the maximum quefrency range. The lower value for this range was chosen, according to [15], in order to reduce the effect of very low quefrency data on the straight line fit.

- **Shimmer local**: a PointProcess object, that represents a sequence of points (or glottal pulses) in the sound, is created through the "*To PointProcess (periodic, cc)*"; the parameters in input 50 and 400 specify the range of pitch periods to be considered. "*Get shimmer (local)*" command allows to calculate the shimmer value, setting a maximum period factor (1.3), that is the largest possible difference between consecutive intervals that is used in the computation of shimmer and maximum amplitude factor (1.6); if the ratio of the durations of two consecutive intervals is greater than these values, this pair of frames are ignored in the computation of shimmer. The values of shortest and longest possible interval are also indicated, respectively to 0.1 ms (default) and to 0.02 (1/50Hz), because intervals longer than this value could be regarded as voiceless stretches and they are ignored for the calculation. The percentage value is obtained by multiplying the result by 100;

- **Shimmer local dB**: extracted in the same way and input parameters used for the shimmer local, with the "*Get shimmer (local dB)*" command;

- **LTAS-slope**: the command "*To Ltas*" computes the Long-Term Average Spectrum (LTAS) of onlyVoice sound, that represents the logarithmic power spectral density as function of frequency, expressed in dB/Hz relative to 2*10-5 Pa. "*Get slope*" is used to obtain the slope value; the algorithm computes the difference between the energy in the frequency low band of [0-1000]Hz and the frequency high band of [1000-10000]Hz of the LTAS, using the energy averaging method (the returned dB value is based on the mean power between the frequency range):

- **LTAS-tilt**: The LTAS between 1 Hz and 10,000 Hz is modelled by a straight line using the "*Compute trend line*" command, and the tilt was calculated based on this line, through "*Get slope*".

- **HNR**: "*To Harmonicity (ac)*" command performs a short-term HNR analysis creates an Harmonicity object, that represents the degree of acoustic periodicity; an HNR of 0 dB means that there is equal energy in the sound harmonics and in the noise. The algorithm performs an acoustic periodicity detection on the basis of the accurate autocorrelation method (chosen to align the algorithm with Matlab scripts) described in [Boersma 1993].

  As input parameters, the following standard values are set: a time step (frame duration) of 0.01s, the pitch floor, that determines the length of the analysis window, of 75 Hz, a silence threshold of 0.1 (frames that contain amplitude relative to the global maximum amplitude below this threshold are considered silent) and a number of periods per window of 4.5.

Calculation of the AVQI

Calculation of the AVQI score based on the equation of Barsties & Maryn (2015) for the AVQI 03.01. [24].

2.7.2 Matlab

Input

Data of each patient are uploaded on Matlab:

- for the AVQI calculation: a concatenation of 3 seconds of the sustained vowel /a/ and a segment of the reading task;
- for the other parameters:
  - 60 seconds of free speech;
  - 3 files, each containing three central seconds of a repetition of the vowel /a/;

Preprocessing

The first part of the script performs a pre-processing that includes:

- **Extraction of the first channel** from each stereo sound;
- **Resampling** of the signals at 44100 Hz;
- **Normalization** of each signal with respect to its maximum in order to have comparable amplitudes.
- Gender of the patient is also identified and different **frequency ranges** are selected for male and female (respectively [75-300] Hz and [100-400] Hz).
- **Selection of non-silent harmonic signal frames with frequency jumps between adjacent frames not greater than half octave.**
  For the removal of silence frames is used a fixed threshold set at half of the RMS value of the entire signal. Each signal is divided into frames of 2048 samples (corresponding to 2048/44100(kSa/s) = 0.046s) and each frame is evaluated: if the RMS value of a frame exceeds the threshold, it is classified as a voiced (non-silence) frame and it is saved for further analysis. Conversely, frames with RMS value below the threshold are considered unvoiced (silence or noise) and are discarded. A second evaluation is performed on frame noise: if the HNR value exceeds zero, indicating that harmonic energy is greater than noise, the frame is marked as harmonic.

The result of this step is an array containing only valid frame (not rejected due to silence or lack of harmonic content).

Parameters extraction

- ***SPL***
  The RMS value of each non-silent frame of 46 ms is calculated through the Matlab function "*rms*" and saved into an array. SPL is computed applying the cited formula 2.11 which allows to convert the rms value in sound pressure (measured at 30 cm from the mouth). The result of this step is an array which contains the SPL value for each frame, from which it is obtained the SPLmean.

- ***HNR***
  To discriminate harmonic from unharmonic signal frames, the evaluation of HNR is computed only on non-silent frames. The HNR value, measured in decibels, using the auto-correlation method described previously, and only frames with an HNR value higher than 0 dB were selected; their values are stored in an array for further analysis. Frames with HNR values below 0 dB are classified as non-harmonic frames and excluded.

- ***CPPS***
  To obtain the CPPS, the signal is resampled to 22050 Hz and CPPS is computed avery 2ms frame length, using a 1024-point Hanning analysis window (of duration of 46 ms). Following the implementation of CPPS described in the software instruction manual MDVP, the Fast Fourier Transform (FFT) algorithm is computed on the windowed signal, obtaining the spectrum amplitude; after this, another FFT algorithm is applied on the log power spectrum in order to obtain the cepstrum domain. The result of this step is the cepstrum of each frame, which shows on the abscissae the quefrency (ms) and the ordinates amplitude in dB. After this, two smoothing steps of the cepstrum are applied:

  1. Time domain: the cepstra of each considered window are smoothed in time using a 14 ms (corresponding to 7 frames); each cepstrum has been replaced by the average of the current frame with the previous three frames and the following three frames;

2. Quefrency domain: a smoothing is computed using a 7-bin averaging window; each cepstral magnitude is replaced by the average of the current bin with the previous and the following three bins.

   A regression line is then calculated on the smoothed cepstrum between 1ms and the maximum quefrency value.

   In this way, it's possible to find the peak in amplitude localised around the fundamental period of the vocal signal: the CPPS is evaluated as the difference between the amplitude peak in the cepstrum and the corresponding value at the same quefrency on the regression line. The research of the peak is conducted only in the range between 3.3 ms and 16.7 ms, that corresponds to the range [60-300] Hz, the typical fundamental frequency range of female and male adults.

- ***Shimmer***
  For extracting the value of local shimmer (%) and shimmer dB, the definitions 2.18 2.17 are applied: voice signal is divided into intervals based on the number of samples per cycle. For each interval, it calculates the peak-to-peak amplitude by finding the difference between the maximum and minimum values within that segment. These peak-to-peak values are then stored in a vector "*vppk*", and the differences between consecutive peak-to-peak values are computed. From these, the script computes the Shimmer Local, expressed as a percentage, which represents the average relative variation in amplitude across cycles, and Shimmer Local dB, which measures the same variation but in decibels, giving a logarithmic representation of the amplitude changes.

- ***Slope***
  Spectral slope refers to the general distribution of energy in the power spectrum. In voiced segments, there is typically more energy concentrated in the lower frequency bands compared to the higher frequency bands. For a normal, modal voice, the energy in the source spectrum usually decreases at a rate of about 12 dB per octave. However, in breathy voices, where the vocal cords stay open for a longer portion of the vibration cycle (extended open phase), the energy drop in the higher frequency bands is even more significant, making the spectral slope steeper.

  To be compatible with the Praat and VOXplot scripts, the slope has been calculated following the Praat definition of the long term average slope reported in table 2.1: the difference between two arbitrary frequency bands ([1-1000]Hz and [1000-10000]Hz) is calculated.

- ***Tilt***
  Spectral tilt has been computed calculating the slope of the regression line through the LTAS. Before computing the slope, the power spectrum is As the slope..

- ***Calculation of the AVQI***
  The AVQI score is determined using the formula provided by Barsties & Maryn (2015) for AVQI version 03.01. [24].

The main differences between MATLAB and Praat lie in the method of silence removal and the fact that MATLAB operates on 46 ms frames, extracting parameters from each individual frame; this approach allows for the calculation of statistical measures for the parameters, such as mean, median, standard deviation, mode, range and others.

2.7.3 VOXPlot

The third software used is VOXPlot (Version: 10.0.22621); Unlike other software which require manual configuration and additional input settings to calculate parameters, VOXplot is fully automatic.

No manually parameter configuration is needed: it is sufficient to upload the two audio signals (the chosen reading passage and the central 3-second sustained vowel /a/) and select the desired analysis language (in this case Italian), which is required to calculate reference values for each parameter.

VOXplot then automatically performs the vocal analysis, calculating all parameters, including AVQI (VOXplot implements the current AVQI v.03.01).

2.8 Vocal analysis

Obtained several objective parameters from the patients' recordings, a vocal analysis was perfomed, to study:

- if there has been any improvement in patients' voice between the start (T0) and the and (T1) of the treatment. The variations between the two temporal observations were calculated as difference: Δ values were calculated for each parameter and index (a positive Δ indicates a voice improvement).

- if there are evident differences between outcomes of the patients treated with the in-clinic or remote rehabilitation. The mean deltas of several parameters are calculated and compared between the two groups of treatment. In the case of free speech, delta values are obtained for each patient according to the formula T1 - T0, and then delta values of each group are averaged; in the case of the sustained vowel /a/, 3 repetitions were available for each time observation, so each parameter includes a mean and a standard deviation. To evaluate the uncertainty of the measured parameters, allowing for a more accurate comparison between telerehabilitation and in-clinic rehabilitation, an uncertainty that considers both inter-patient variability (differences between patients of the same group) and intra-patient variability (fluctuations within repeated vowels of the same patient) was calculated. Following the approach introduced in [25], the uncertainty for the telerehabilitation group can be computed following these steps:

    1. Calculation of the mean value $x_{i,\text{tele}}$ and the standard deviation $u(x_{i,\text{tele}})$ for each patient:

    $$x_{i,\text{tele}} = \frac{1}{n}\sum_{j=1}^{n} x_{i,j}$$

    where n is the number of repeated measurements (e.g., across 3 different vowels) for each patient i.

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

    $$M_{\text{tele}} = \frac{1}{N}\sum_{i=1}^{N} x_{i,\text{tele}}$$

where N is the number of patients that followed telerehabilitation, in this case $N = 10$.

3. Calculation of the type A standard uncertainty $u_A(M_{\text{tele}})$, related to the dispersion of each subject with respect to the class mean-value:

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

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

$$u_B(M_{\text{tele}}) = \frac{1}{N}\sqrt{u^2(x_{1,\text{tele}}) + \cdots + u^2(x_{N,\text{tele}})}$$

   This component accounts for the intra-patient variability, i.e., the consistency of repeated measurements for each patient.

5. Calculation of the combined standard uncertainty $u(M_{\text{tele}})$, that combines the effect of both inter and intra-patient variability:

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

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

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

   where $k = 2$ is the coverage factor, providing a 95% confidence interval.

7. Repetition of steps for the in-clinic rehabilitation group: After following the same steps for the in-clinic rehabilitation group, the two uncertainties $U(M_{\text{tele}})$ and $U(M_{\text{pres}})$ can be compared to assess the consistency and reliability of the two different therapeutic approaches.

- if there are any differences between the results obtained using different software applications, previously described, or if they can be considered comparable.

2.9 Patient autovalutation

During the study the subjective perception of the patients at the start (T0) and at the end (T1) of the therapy was also evaluated: the Voice Handicap Index (VHI) is an auto-evaluation scale commonly used to measure the perceived vocal handicap and monitor improvements following treatment.
A questionnaire was given to each subject, where he has to evaluate his vocal handicap degree, through 30 questions divided in three groups, regarding three domains:

- functional: the impact of vocal disorders on daily activities;
- emotional: the patient's affective response to the voice disorder;
- physical: self-perception of features of laryngeal discomfort and voice output characteristics.

For each statement a score has to be assigned, from 0 (never) to 4 (always) [26]. Summing all the answers, a global score is obtained, ranging from 0 to 120: 0 value is correlated with a normal voice, and a high number indicates greater severity of voice problems.

	Voice alteration			
	Normal	Slight	Moderate	Severe
VHI	0	1-40	41-80	81-120

The positive results of the patient's self-assessment measure at the end of the rehabilitation process would strengthen the hypothesis of the effectiveness and efficiency of treatment.

Chapter 3

Results

In this chapter, the results obtained from the study are presented and discussed.

First, AVQI values extracted using Praat software and Warning Score are used to assess the effectiveness of the rehabilitation process. This involves comparing and analyzing vocal parameters at two distinct time points: T0, which represents the baseline measurements taken before the rehabilitation program begins, and T1, which corresponds to the measurements collected at the end of the rehabilitation. Furthermore, this analysis explores the comparative efficacy of tele-rehabilitation versus in-clinic rehabilitation, including also an analysis on the SPL values calculated on free speech data, and four parameters (jitter, shimmer, HNR, CPPS) extracted from the sustained vowels.

The Warning Score values calculated on data provided by both Matlab and Vocal Holter are presented to examine any differences observed between the two microphones (in-air and contact) used.

An autovalutation of patients, called VHI, is also considered to quantify self-assessment of voice quality and observe how patients feel before and after the therapy.

Additionally, a comparison between different software tools, including VOXPlot, Matlab and Praat is performed to highlight any discrepancies between them.

3.1 Validation of the effectiveness of rehabilitation process

3.1.1 AVQI

The Acoustic Voice Quality Index, as described in paragraph 2.2.1, ranges between 0 and 10 and depends on several parameters: shimmer, shimmer dB, CPPS, HNR, spectral slope and tilt. These parameters are extracted by the concatenation of 3 central seconds of sustain vowel /a/ and a segment of reading task. To ensure the accuracy of the AVQI calculation, each patient was asked to sustain the vowel sound three times; acoustic parameters were calculated for each repetition, and these parameters were the averaged to provide a more consistent and representative measure. The AVQI values showed in these results are a mean values of the three AVQI obtained from the concatenation of the continuous speech with each vowel repetition.

First, the extraction of the parameters and the calculation of the AVQI index were performed in Praat. The three-second sustained vowel segments and the pass of reading were renamed "sv" and "cs", respectively; to execute the analysis, the script developed by Maryn, reported in Appendix A, is utilized. It automates the extraction of the relevant acoustic parameters and computes the AVQI 03.01 index.

The comparison was made between the parameters extracted from the voice signal before starting rehabilitation therapy (T0) and after the completion of the rehabilitation process (T1).
Additionally, differences in outcomes between patients undergoing tele-rehabilitation and those receiving in-clinic rehabilitation were assessed, highlighting variations in the effectiveness of these two approaches: AVQI value is used as one of the treatment outcome measures.

As described in the AVQI formula, each parameter contributes with a specific weigth to the final index and represents a different aspect of voice quality.
For an improvement in the AVQI:

- the **shimmer** should decrease (negative Δ): this value measure the amplitude perturbation. A lower shimmer indicates a more stable vocal signal, with less amplitude variability.

- the **HNR** should increase (positive Δ): an higher HNR indicates a clearer and less noisy signal.
- the **CPPS** should increase (positive Δ): an healthy voice presents a greater periodicity of the signal, an high cepstral peak and so an higher CPPS.
- the **slope** and the **tilt** should increase (positive Δ): these spectral parameters refer to the global distribution in the power spectrum (LTAS) of a vocal signal. In healthy voices, the power spectrum shows a significant concentration of energy in the lower frequency bands, with a gradual decrease in higher frequencies. In breathy phonation, an extended open phase during vocal fold vibration allows more airflow and increases turbolence: the spectral energy is concentrated in lower frequencies, causing a flatter slope and a decreased tilt.

Tables 3.1 and 3.2 present the parameters obtained respectively at T0 and T1 for all the patients; the first ten patients are those who went regularly in clinic, while the last ten, highlighted in light gray, represent the patients who were treated with remote rehabilitation. The reported AVQI for each patient represents the mean value of the three AVQI obtained using three different sustain vowel samples.

In table 3.2, AVQI improvements from T0 to T1 are highlighted in green, and quasi-stable conditions are marked in yellow.

Table 3.1: Praat AVQI results at **T0**;
patients treated with tele-LSVT are reported in gray;

ID	$CPPS_{mean}$ (dB)	HNR (dB)	Shimmer (%)	Shimmer (dB)	Slope (dB)	Tilt (dB)	AVQI
1	9.20	11.82	13.59	1.26	-24.87	-7.13	6.42
3	6.58	12.43	13.18	1.26	-29.21	-3.64	8.61
5	12.17	13.70	10.10	0.98	-20.13	-8.07	4.36
6	9.71	10.13	12.66	1.20	-18.42	-9.31	5.73
8	9.27	10.68	14.89	1.36	-23.39	-9.02	6.07
11	6.05	10.10	15.90	1.41	-28.84	-8.79	7.66
13	8.50	11.73	14.25	1.31	-26.26	-6.64	6.94
15	9.19	12.11	10.99	1.11	-23.45	-4.87	6.95
19	11.89	10.58	9.84	1.07	-11.76	-7.31	5.28
20	8.52	11.98	12.44	1.18	-27.72	-7.58	6.48
2	10.79	13.49	10.24	1.06	-20.53	-6.73	5.62
4	10.17	13.50	11.04	1.04	-26.40	-8.40	5.20
7	9.47	11.01	13.09	1.20	-25.76	-8.24	5.87
9	10.69	12.14	10.31	0.99	-19.90	-8.19	5.14
10	11.33	12.86	11.66	1.07	-23.74	-8.33	4.74
12	12.23	12.55	11.33	1.10	-19.10	-9.49	4.21
14	4.37	7.99	18.68	1.60	-29.52	-7.62	9.10
16	10.99	12.55	10.97	1.02	-17.44	-8.23	5.05
17	9.87	10.55	10.23	0.91	-23.58	-6.04	5.84
21	10.44	13.22	9.54	0.93	-21.81	-7.22	5.37

Table 3.2: Praat AVQI results at **T1**;
patients treated with tele-LSVT are reported in gray;

ID	CPPS$_{mean}$ (dB)	HNR (dB)	Shimmer (%)	Shimmer (dB)	Slope (dB)	Tilt (dB)	AVQI
1	10.84	13.63	11.17	1.04	-25.56	-8.52	4.82
3	8.33	12.87	10.15	1.03	-25.30	-4.62	7.26
5	13.29	14.81	6.60	0.72	-17.63	-7.34	3.72
6	11.75	11.91	10.59	1.09	-16.54	-10.25	4.39
8	11.59	14.07	10.72	1.09	-15.04	-9.12	4.75
11	6.50	11.71	12.87	1.19	-27.33	-8.20	7.34
13	8.76	12.80	12.89	1.25	-26.18	-7.07	6.66
15	11.25	13.15	8.26	0.90	-17.43	-4.94	5.74
19	10.94	10.90	12.37	1.13	-17.64	-7.37	5.48
20	11.15	13.01	10.12	1.01	-25.18	-7.98	4.88
2	12.13	15.77	9.65	0.99	-11.40	-6.09	5.22
4	10.67	15.62	8.38	0.92	-26.86	-8.89	4.70
7	11.52	12.67	10.73	0.98	-24.87	-8.47	4.42
9	9.62	10.82	11.30	1.04	-19.59	-8.20	5.74
10	12.79	12.93	7.26	0.92	-12.46	-7.72	4.53
12	14.61	16.39	6.58	0.76	-12.03	-8.41	3.01
14	6.33	9.45	16.22	1.51	-25.77	-6.88	8.39
16	11.42	13.73	8.53	0.87	-15.42	-8.91	4.53
17	10.10	9.21	13.68	1.18	-22.28	-6.56	6.01
21	11.94	14.39	8.69	0.85	-20.43	-8.26	4.24

Overall, there is a noticeable improvement from T0 to T1 for most of the patients; in fact, excluding patients 9, 16, 6 and 19, where the conditions were remained quasi-stable, there is a general decrease of the AVQI values.

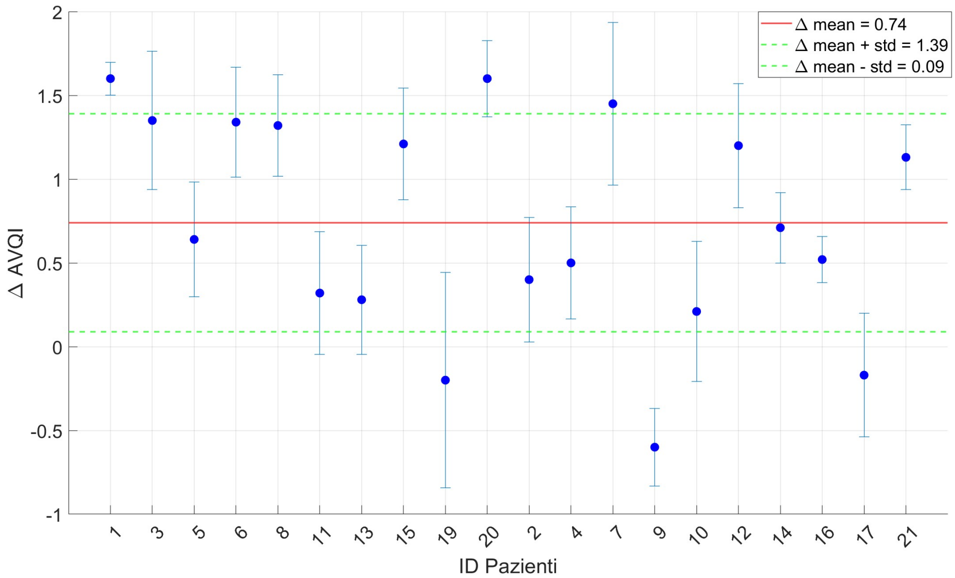

Figure 3.1: Delta AVQI in Praat

Figure 3.1 shows the Δ values for each patient, calculated as difference between values in T0 and values in T1, including also the variability intra-patients across the three AVQI values (calculated on the three sustained vowels).
Overall, the results reveal a positive mean Δ obtained ($\Delta AVQImean = 0.74$), suggesting a general improvement in vocal quality among the patients, and thereby reflecting the effectiveness of the rehabilitation programs.

To compare the two modalities of treatment, the outcomes of in-clinic versus tele-rehabilitation protocol are plotted in figure 3.2, that shows the average ΔAVQI for both groups. With the AVQImean values is also figured the respective uncertainity, that combines both the variability of the three vowel repetitions and the type of therapy followed.
In both treatments an improvement of vocal quality can be observed, resulted as ΔAVQI of 0.9 ± 0.15 for the in-clinic group and 0.6 ± 0.18 for the tele-rehabilitation group; although the in-clinic group demonstrated slightly greater improvement, the

overlapping uncertainty bars indicate that the two treatments can be considered equivalent in their effectiveness.

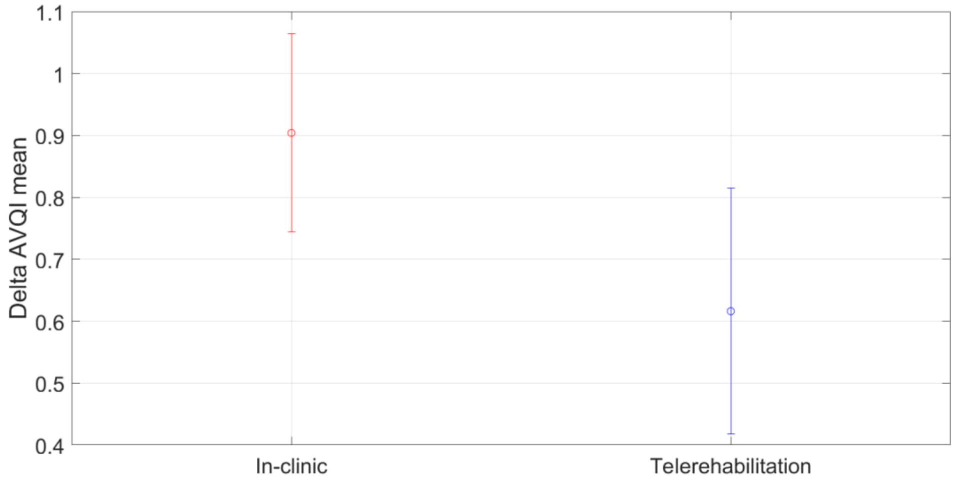

Figure 3.2: Delta AVQI in Praat: tele-rehabilitation VS in-clinic

3.1.2 SPL on free speech

In this chapter is evaluated the impact of the LSVT-Loud on vocal loudness, misured with Sound Pressure Level (SPL), introduced in paragraph 2.2.1. Individuals with multiple sclerosis often presents a weaker voice (lower SPL) than normal: LSVT-Loud treatment maximizes phonatoy efficiency by enhancing vocal fold adduction and optimizing laryngeal muscle activation and control through high-effort, loud phonation exercises.

SPL was calculated for the monologue, a free speech chosen by the patient, as his morning routine, during 60 seconds. The SPL mean values reported are extracted using Praat and are referred to a distance of 30cm from the mouth, the distance at which is the in-air microphone is placed during the recordings.

To convert it to SPL at 1m, often used in literature, formula 2.13 is computed.

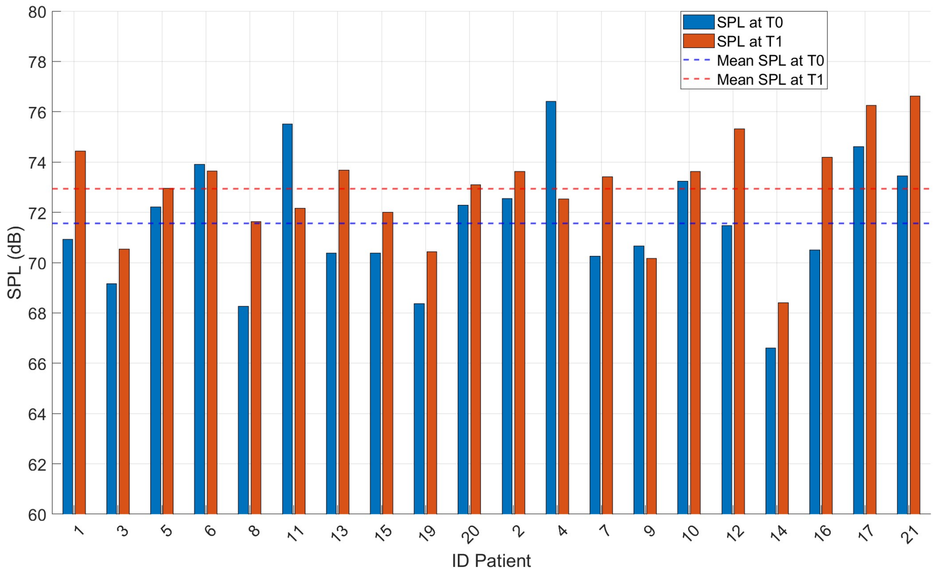

Figure 3.3: SPL values at 30cm at T0 and at T1, calculated in Praat

By examining the table 3.3 and the figure 3.3 it is evident that there has been an overall improvement in SPL values from T0 to T1: specifically, the mean SPL increased from 71.55 to 72.94.

However, for two patients, 6 and 9, the mean SPL remained mostly stable, while decreased in patients 4 and 11. This decrease can be attributed to the fact that the

Table 3.3: SPL for monologue at T0 and T1
patients treated with tele-LSVT are reported in grey;

	SPL at T0 (dB)		SPL at T1 (dB)	
ID	**30 cm**	**1m**	**30 cm**	**1 m**
1	70.93	60.47	74.44	63.98
3	69.16	58.70	70.54	60.08
5	72.21	61.75	72.95	62.49
6	73.91	63.45	73.65	63.19
8	68.26	57.80	71.64	61.18
11	75.51	65.05	72.16	61.70
13	70.38	59.92	73.67	63.21
15	70.38	59.92	72.01	61.55
19	68.36	57.90	70.43	59.97
20	72.29	61.83	73.09	62.63
2	72.54	62.08	73.62	63.16
4	76.41	65.95	72.53	62.07
7	70.26	59.80	73.41	62.95
9	70.67	60.21	70.17	59.71
10	73.23	62.77	73.62	63.16
12	71.48	61.02	75.32	64.86
14	66.60	56.14	68.40	57.94
16	70.50	60.04	74.18	63.72
17	74.62	64.16	76.25	65.79
21	73.44	62.98	76.62	66.16

recording of the monologue of these patients took place in a different room than at T0: differences in room size or in background noise could have influenced the patients to speak at a lower volume during T1. For record the monologue, patients were asked to speak freely using a comfortable tone, which may vary depending on external noise. Moreover, there may have been slight inaccuracies regarding the 30 cm distance between the microphone and the mouth: as patients speak, they may move their heads slightly, and it can be challenging to immobilize them, especially considering that many of them have motor diseases. Even small variations in the microphone's position, by just a few centimeters, could negatively impact the results. Excluding these two outliers, the SPLmean would increase from 71.07 dB to 73.00 dB.

Separating the two groups of patients treated with different types of rehabilitation, Δ SPL was calculated as the mean of the differences between SPL in T1 and SPL at T0; a positive delta is correlated with an improvement in voice intensity.

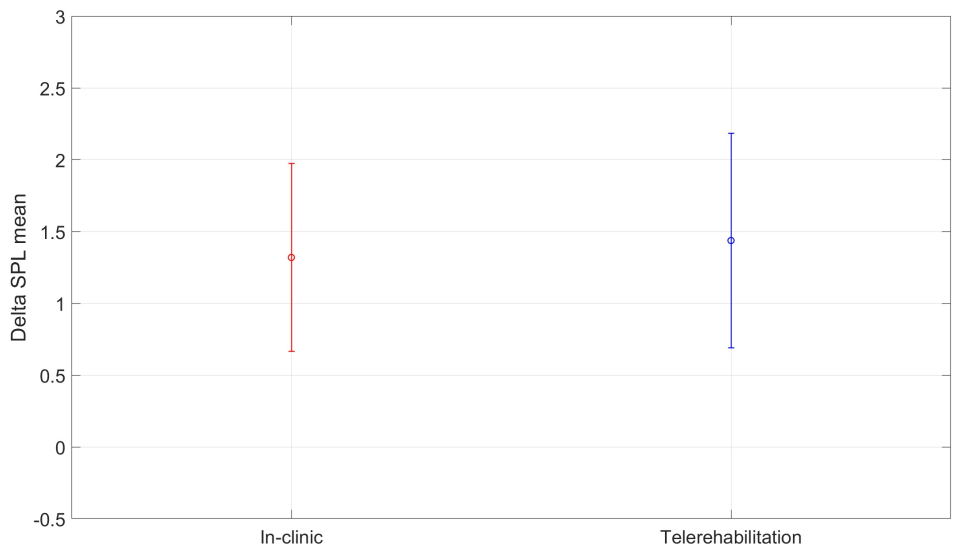

Figure 3.4: SPL values at 30cm at T0 and at T1, calculated in Praat

As can be seen in figure 3.4 there is no evident difference between the Δ of the two groups of rehabilitation: in-clinic and remote treatment can be considered equivalent in term of improvement of SPL.

3.1.3 Parameters on sustained vowels

To evaluate the effectiveness of the LSVT-Loud, the following parameters extracted by the sustained vowels are also taken into account: Jitter (%), Shimmer (%), HNR (dB) and CPPS (dB).
Parameters shown in tables 3.4, 3.5 are extracted using a Praat script, for T0 and T1 and averaged on the three repetitions on the entire vowel /a/ for each patient.

Table 3.4: Vowel's parameter at **T0**;
patients treated with tele-LSVT are reported in gray;

ID	Jitter (%)	Shimmer (%)	HNR (dB)	CPPS (dB)
1	0.67	12.26	11.41	10.61
3	1.78	7.37	16.40	7.85
5	0.61	7.92	13.93	14.49
6	1.06	8.83	11.90	11.00
8	0.93	9.07	8.89	10.09
11	1.19	13.09	8.12	7.32
13	0.97	15.28	9.15	9.68
15	0.79	7.47	14.55	10.91
19	0.53	7.55	10.09	16.94
20	0.96	10.05	12.67	11.85
2	0.49	8.59	14.18	13.92
4	0.45	6.80	12.93	12.50
7	0.50	12.60	12.29	11.32
9	0.77	9.25	13.76	12.38
10	0.35	8.44	13.84	13.73
12	0.35	6.79	14.59	15.71
14	1.78	8.60	9.11	5.94
16	0.47	9.56	16.10	13.51
17	0.75	6.08	11.28	11.91
21	0.60	7.58	13.75	12.21

Table 3.5: Vowel's parameter at **T1**;
patients treated with tele-LSVT are reported in gray;

ID	Jitter (%)	Shimmer (%)	HNR (dB)	CPPS (dB)
1	0.52	8.28	14.96	12.92
3	1.09	5.85	15.62	10.49
5	0.47	4.13	16.12	15.59
6	0.60	6.15	11.76	13.86
8	0.27	5.05	17.13	15.69
11	1.36	9.01	11.57	8.01
13	0.73	15.12	10.61	10.17
15	0.30	5.11	14.83	14.52
19	0.73	9.19	11.55	14.43
20	0.63	7.58	12.56	14.70
2	0.25	5.05	17.18	17.09
4	0.26	4.35	19.12	15.08
7	0.34	8.81	13.97	12.58
9	1.11	8.84	10.02	10.36
10	0.22	4.26	17.59	19.34
12	0.36	3.00	17.48	20.14
14	0.78	11.74	8.27	8.82
16	0.47	6.95	14.29	15.44
17	0.51	8.93	10.23	13.19
21	0.56	5.59	16.32	13.76

As it can be seen in figure 3.5, there is a general improvement across all the patients in the four extracted parameters from T0 to T1:

- Jitter (%) and Shimmer (%) mean values at T1 decreased respectively of 27.5% ($\Delta_{\text{jitter}} = -0.22\%$) and 21.94% ($\Delta_{\text{shimmer}}$ = -2.01%)from T0;

- HNR (dB) and CPPS (%) mean values at T1 increased respectively of 12.93% (Δ_{HNR}= +1.61 dB) and 18.14% (Δ_{CPPS} = +2.12 dB) from T0.

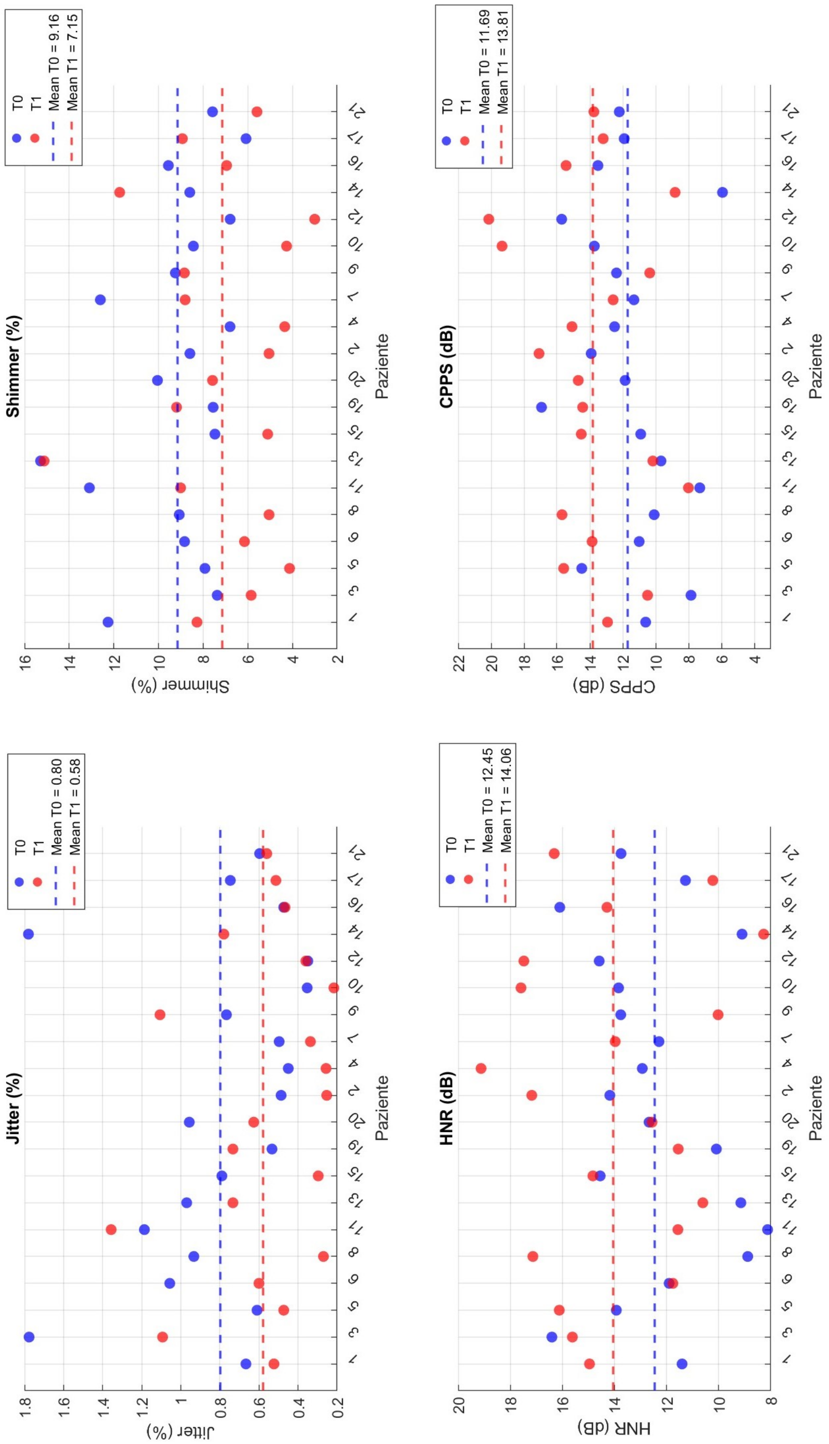

Figure 3.5: Comparison of vowel parameters at T0 and T1

In figure 3.5 deltas of each parameter are shown: positive deltas are correlated to an improvement of voice quality. For this reason, for the jitter and the shimmer the difference between T0 and T1 was computed (lower values indicate healthier voices), while for HNR and CPPS the delta values were calculated for T1-T0 (higher values indicate healthier voices). For the comparison between the two types of rehabilitation, the mean value of each parameter across the patients of the same group was computed, with the combinated uncertain that takes into account both intra and inter-patient variability.

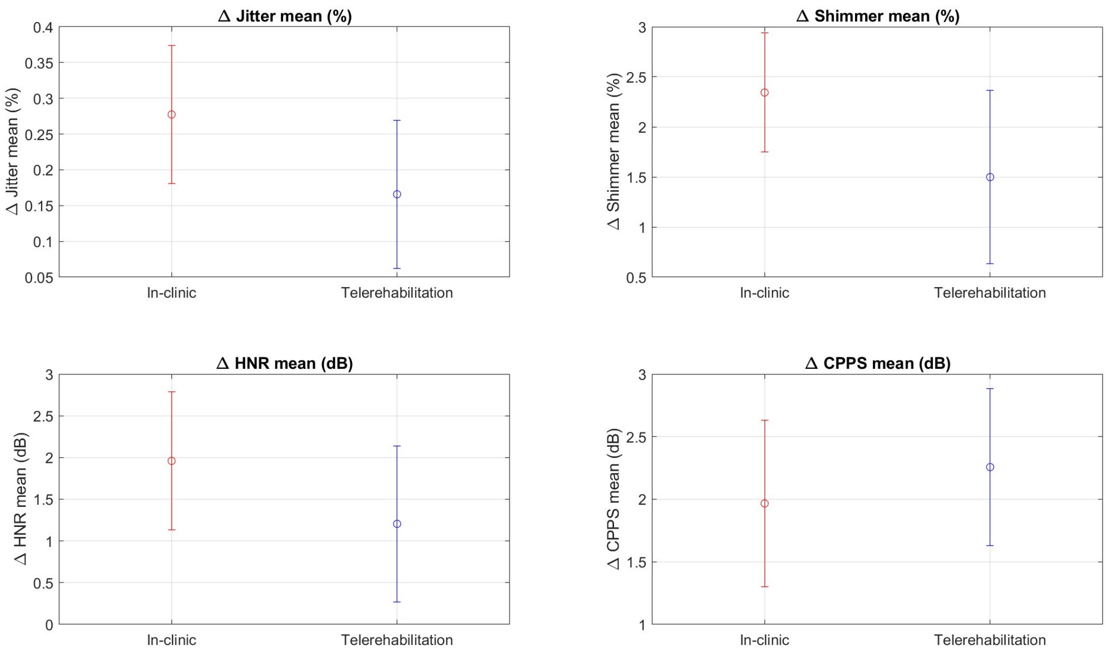

Figure 3.6: Comparison of therapies (in-clinic and telerehabilitation) in terms of Δ

The comparison between the two rehabilitation groups, based on the sustained vowels, shows that:

- **jitter**: $\Delta_{\text{jitter}} = (0.28 \pm 0.10)\%$ for the in-clinic group and $\Delta_{\text{jitter}} = (0.17 \pm 0.10)\%$ for the telerehabilitation group;
- **shimmer**:$\Delta_{\text{shimmer}} = (2.34 \pm 0.59)\%$ for the in-clinic group and $\Delta_{\text{shimmer}} = (1.50 \pm 0.87)\%$ for the telerehabilitation group;

- **HNR**: $\Delta_{\text{HNR}} = (1.96 \pm 0.30)$ dB, for the in-clinic group and $\Delta_{\text{HNR}} = (1.21 \pm 0.94)$ dB for the telerehabilitation group;
- **CPPS**: $\Delta_{\text{CPPS}} = (1.97 \pm 0.67)$ dB for the in-clinic group and $\Delta_{\text{CPPS}} = (2.26 \pm 0.63)$ dB for the telerehabilitation group.

The improvements in these parameters indicate that both rehabilitation approaches yield comparable results: even if in jitter, shimmer and HNR the in-clinic Δ mean are higher than those for tele-rehabilitation group, the combined uncertainty of the Δ are overlapped.

Also for the sustained vowel's analysis, the two types of rehabilitation can be considered equivalent.

3.1.4 Warning Score and comparison between different microphone

The Warning Score, as described in paragraph 2.2.2, is a new index used to assess the state of vocal health of patients.
The score is assigned taking in account different parameters extracted from the repetitions of the sustained vowel /a/, as local jitter (%), local shimmer (%), median and standard deviation of the CPPS (dB).

Each parameter is evaluated against predefined thresholds that have been established through clinical research, and contributes to the final Warning Score with:

- +1 : if the parameter falls within pathological ranges, suggesting potential vocal issues;

- -1 : if the parameter is within the "healthy range", indicating normal vocal function;

- 0: the parameter doesn't give contribution to the calculation if it's in an indeterminate range, considered not reliable.

These cut-off values allow for the classification of vocal quality, helping to distinguish between healthy and pathological voices.
The Warning Score, represented by the sum of these contributions, ranges from -4 to +4, with positive values correlated with pathological voices, while negative values are associated with healthy ones.

The parameters extracted by the Vocal Holter device (described in paragraph 2.5.2) for the sustained vowel /a/ were processed in order to evaluate a warning score for each involved subject for T0 and T1. The thresholds used for each parameter were assigned following the rules implemented in a previous study [21].
The reported values are the average of the parameters obtained from the three repetitions of the sustained vowel.

Table 3.6: Warning Score at T0 - VOCAL HOLTER
patients treated with tele-LSVT are reported in gray;

ID	Jitter (%)	Shimmer (%)	CPPSmedian (dB)	CPPSstd (dB)	WS
1	0.78	2.62	16.67	2.76	4
3	1.61	6.03	10.80	1.94	4
5	0.65	2.55	16.80	1.57	3
6	1.28	5.82	17.60	3.20	4
8	1.05	3.44	13.47	1.90	4
11	1.38	8.53	13.00	1.79	4
13	0.70	2.07	15.40	1.20	1
15	0.51	1.51	15.63	1.38	2
19	0.47	2.18	17.20	1.47	2
20	2.72	6.69	15.53	3.91	4
2	0.23	1.10	18.07	1.05	-2
4	0.45	3.66	16.73	1.48	4
7	0.42	5.05	15.30	1.15	2
9	1.17	4.73	16.97	1.85	4
10	0.33	2.38	18.70	1.65	1
12	0.64	1.90	16.90	1.72	2
14	1.69	11.59	12.00	3.18	4
16	0.26	1.89	15.07	1.16	-1
17	0.64	5.19	17.17	1.66	4
21	0.40	2.40	15.36	1.09	1

Table 3.7: Warning Score at T0 -VOCAL HOLTER
patients treated with tele-LSVT are reported in gray;

ID	Jitter (%)	Shimmer (%)	CPPSmedian (dB)	CPPSstd (dB)	WS
1	0.59	3.13	17.40	1.81	4
3	0.55	3.56	14.80	1.87	4
5	0.38	1.61	17.83	1.74	1
6	0.56	4.02	18.67	1.27	2
8	0.20	1.57	17.77	0.90	-1
11	0.99	5.13	13.43	1.55	4
13	0.48	2.07	15.53	1.02	1
15	0.26	1.23	17.20	0.96	-1
19	0.82	2.61	15.53	1.45	4
20	1.01	3.51	18.40	2.60	3
2	0.18	1.09	18.47	0.86	-3
4	0.22	1.86	18.60	1.36	-1
7	0.25	11.55	17.20	1.38	2
9	1.18	10.00	17.43	0.88	2
10	0.21	1.74	18.60	1.50	-1
12	0.32	8.04	18.33	1.95	2
14	1.12	6.37	13.80	1.90	4
16	1.09	3.27	15.80	1.78	4
17	0.46	4.35	18.76	1.82	3
21	0.40	2.72	15.96	1.16	2

The delta value of Warning Score was calculated for each patient, as $\Delta_{\text{WS}} = WS(T0) - WS(T1)$; a positive delta of the index indicates an improvement in voice quality.
Figure 3.7, where the two groups of patients are separated, highlights a positive variation trend in both the in-clinic and telerehabilitation groups in terms of WS. Data show that most patients experienced either a reduction or no significant change in their warning score ($\Delta_{\text{WS}} \geq 0$) and the two treatments showed similar patterns. In each group there is an outlier (patient 19 and patient 16), which presents a substantial worsening of the warning score, contrasting the general pattern.

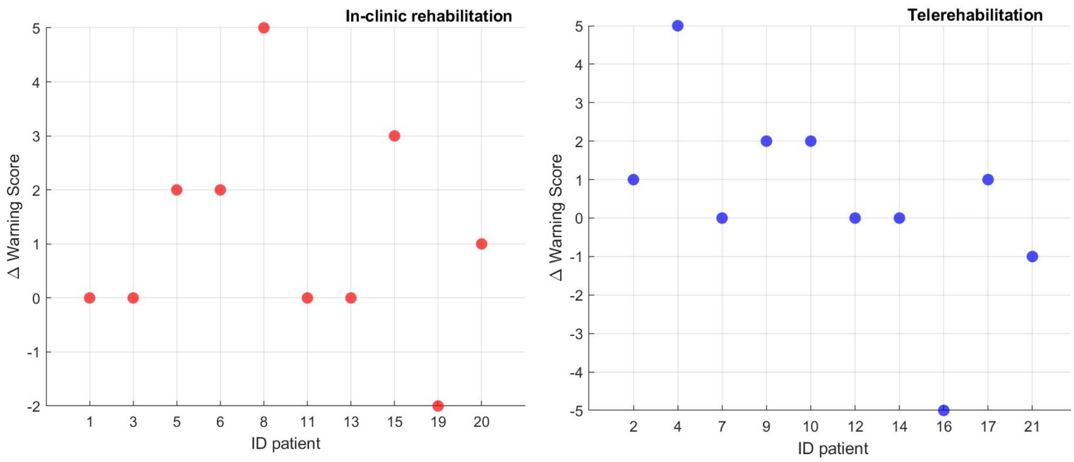

Figure 3.7: Delta values for warning score

A further analysis has been conducted focusing on the type of microphone used: in order to validate the Vocal Holter device for the computation of the Warning Score, a comparison between the VH and the in-air microphone has been performed.

In the first instance, the acoustic parameters were extracted by the contact microphone integrated into the Vocal Holter; in the second case, parameters were extracted, using Matlab (Praat and VOXPlot software do not provide the median and the standard deviation of the CPPS), from the recordings acquired by the in-air microphone.

Identical values between the two microphones were not expected as the input signals they capture are different: the in-air microphone acquires sound pressure waves in the air, modulated by the vocal tract, while the VH measures the vibration induced by the vocal folds at the neck, which acts as a natural low-pass filter.

As shown in table 3.8, the greater difference between the two microphones is observed in the shimmer values:

- jitter, CPPSmedian and CPPSstd differ respectively of 3,1%, 9.5% and 13.7%;

- VH's shimmer values are, on average, 44,5% lower than those extracted using the in-air microphone.

Table 3.8: Comparison between Vocal Holter and in-air microphone
for each parameter the mean and the standard deviation are reported

		JITTER (%)		**SHIMMER (%)**		**CPPSmedian (dB)**		**CPPSstd (dB)**	
		mean	***std***	***mean***	***std***	***mean***	***std***	***mean***	***std***
T0	VH	0.87	0.62	4.07	2.68	15.72	2.05	1.85	0.79
	IN-AIR	0.83	0.49	8.03	3.19	14.33	2.29	2.01	0.51
T1	VH	0.56	0.35	3.97	2.93	16.98	1.67	1.49	0.45
	IN-AIR	0.57	0.43	6.59	2.51	15.1	1.85	1.77	0.42

To assess the variability in the estimation of these parameters between the two devices, the differences between the parameters extracted with the microphone in air and the parameters stored inside the DAP unit of the VH are reported($\Delta = MIC - VH$). To avoid redundancy, only the data at T0 are presented, as T1 graphs showed similar results.

Deltas for each subject are presented; in order to provide a clearer understanding of the results, are also reported the mean delta and the standard error (calculated by dividing the standard deviation by the square root of the total number of subjects). To validate the use of the Vocal Holter, the delta should be nearer as possible to the zero value.
For local shimmer measure (Fig. 3.8) significant differences between the two microphones are present. A mean value of delta shimmer of 3.96% at T0 is found, indicating that the shimmer values obtained with the microphone in air are consistently higher, on average, than those measured with the VH.

Even after excluding the values of patient 4 and 14, that can be considered outliers likely due to a malfunction of the Vocal Holter, the Δ between the two microphones remains at 3.75%. This value is still considered excessively high, comparing it with the threshold cited above to classify the local shimmer.

These results are reliable and can be explained by the distinct nature of the input signals captured by the two microphones.

The contact microphone of the VH acquires the mechanical vibrations from the neck and is less sensitive to high-frequency components, because the tissues of the neck act as a low-pass filter, smoothing out rapid fluctuations in amplitude which directly affect shimmer (representing the stability in amplitude of the signal). In

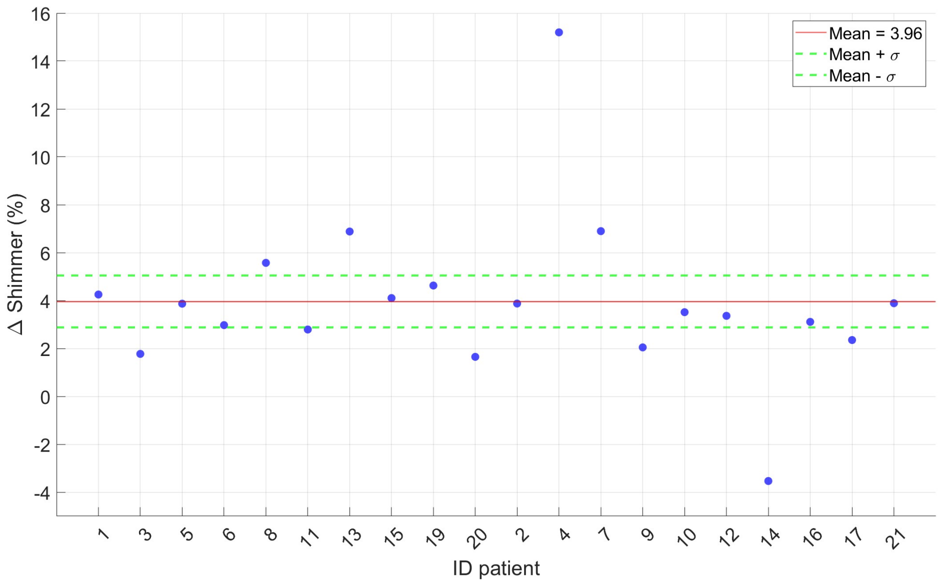

Figure 3.8: Delta Shimmer as difference AIRMIC - VH

contrast, the in-air microphone acquires a more complex signal, shaped by the vocal tract.

Moreover, the in-air wave is also influenced by other sound sources in the environment, unrelated to the vocal signal; the VH is insulated from the environmental noise, resulting in a more stable analysis.

These results are supported by the values obtained for the jitter (fig. 3.9): this parameter, which indicates the fundamental frequency perturbation, is less affected by background noise sources and shows a lower Δ value between the two microphones.

A mean Δ jitter of -0.038% was calculated, indicating that the jitter values extracted by the two microphones are very similar.

This analysis is also conducted on CPPSmedian and CPPSstd (fig. 3.10).

The Δ CPPS standard deviation values is of +0.16 dB; excluding patient 4, which, as seen in shimmer values, can be considered an outlier due to a VH malfunction, possibly caused by improper placement of the contact microphone, was calculated Δ value for the CPPSstd of 0.06 dB, considered no significant. For the CPPS median, the Δ value obtained is of -1.39 dB: this negative value indicates that the value of CPPS median measured in air is lower than that provided by VH. This

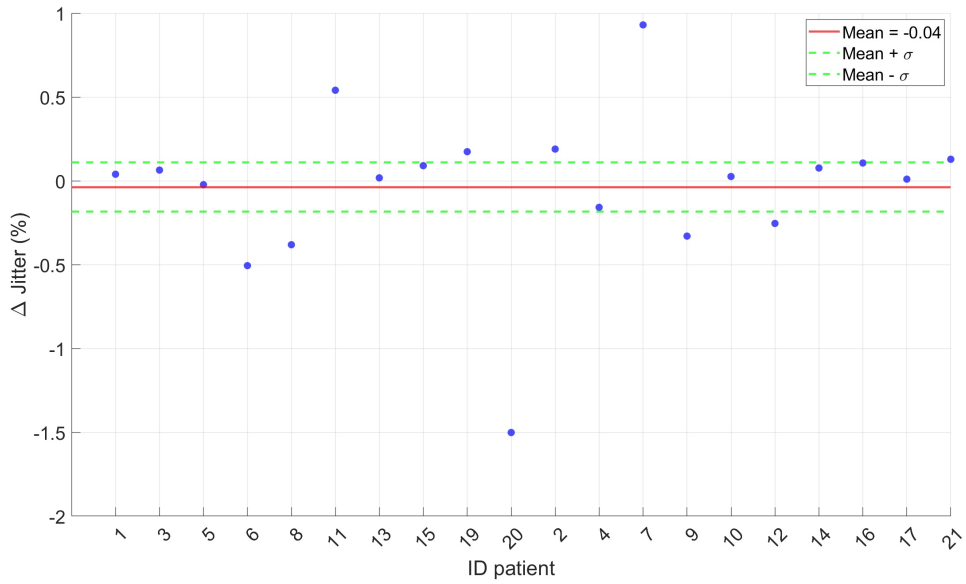

Figure 3.9: Delta Jitter as difference AIRMIC - VH

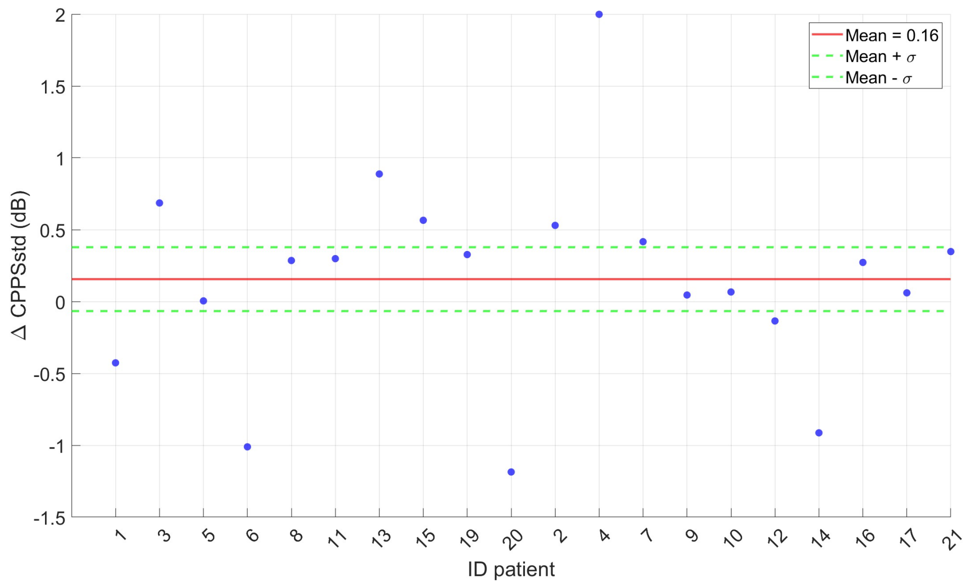

Figure 3.10: Delta CPPS std as difference AIRMIC - VH

difference between the CPPS median values, as observed in [**castellana2018**], can be attributed to the fact that CPPS distributions vary with the characteristics of the measurement chain, i.e. inner noise floor and bandwidth of the devices: while the contact microphone embedded in the Vocal Holter has a frequency content of approximately 3.5 kHz, the in-air microphone has 10 kHz.

Because of these differences between the two microphones, the WS values obtained with the in-air microphone differ from those extracted from the DAP unit of the Vocal Holter. In fact, the thresholds for assigning the points to obtain the WS were assigned using the VH parameters.

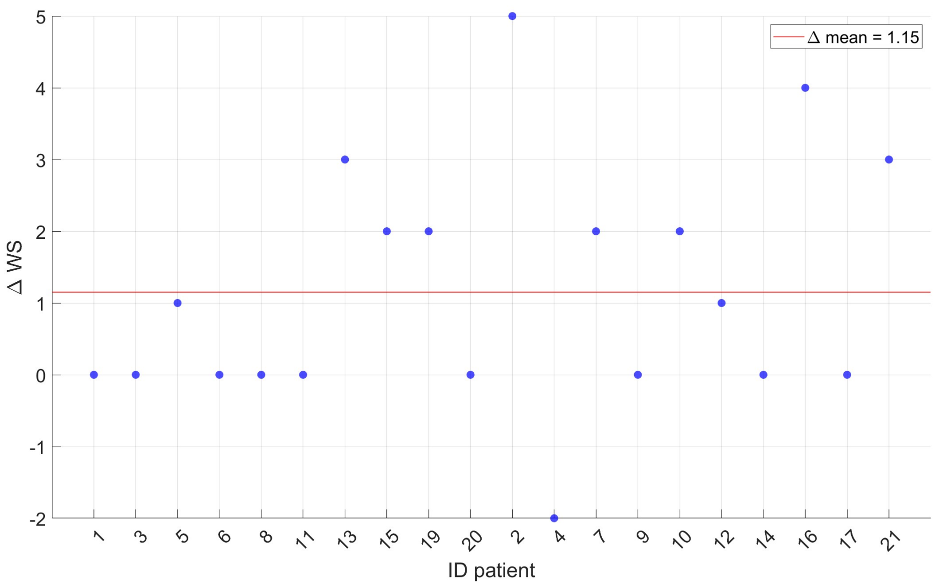

Figure 3.11: Delta WS as difference AIRMIC - VH

Figure 3.11 shows the Δ_{WS} values, which are calculated as the difference between the warning scores obtained from the air microphone and those from the Vocal Holter ($\Delta_{\mathrm{WS}} = \mathrm{WS}_{\mathrm{mic}} - \mathrm{WS}_{\mathrm{VH}}$). As seen in the figure, with the exception of more severe cases where the Δ_{WS} is zero (because both devices assign a high warning score of 4), the air microphone consistently tends to assign a higher warning score compared to the vocal holter. This discrepancy is due to the fact that, as observed in previous results, particularly for shimmer, the air microphone returns higher values, often exceeding the threshold: these higher values indicate a more pathological voice, which results in a +1 contribution to the WS. On the other hand, the same parameters extracted from the VH tend to be lower, frequently falling within the unreliable range (contributing 0 to the WS) or the healthy voice range (contributing -1), lowering the final WS.

In future, it would be useful to establish new thresholds to assign the WS using the parameters extracted from the in-air microphone, or try to change the algorithm to make them comparable.

3.2 VHI

After the objective assessment of effectiveness therapy, the vocal handicap index (VHI) is used to quantify self-assessment of voice quality and measure psycho-social effects of voice disorders. An higher value corresponds to a severe vocal alteration and a greater impact of voice problems, while a lower VHI indicates a better patients' self-perception.

Table 3.9: VHI values at T0 and T1
patients treated with tele-LSVT are reported in grey;

ID	VHI_T0	VHI_T1	ID	VHI_T0	VHI_T1
1	53	26	**2**	69	62
3	82	85	**4**	19	19
5	74	67	**7**	22	7
6	55	10	**9**	58	36
8	57	49	**10**	24	11
11	78	64	**12**	16	12
13	30	25	**14**	38	47
15	18	15	**16**	13	9
19	34	41	**17**	33	18
20	21	21	**21**	37	15

Table 3.9 presents a comparison of VHI scores for patients at T0 and at T1, providing a clear indication of the therapy's impact.

Significant reductions in VHI are evident for many patients, reflecting significant improvements in vocal function and quality of life. However, some cases, as VHI of patient 4 and patient 20 remains stable, and in two cases (patients 14 and 19) VHI increases from T0 to T1. Despite these outliers, the overall trend in the table supports the conclusion that voice therapy leads to a reduction in VHI scores, which supports the effectiveness of rehabilitation programs. Additionally, the results highlight the comparable effectiveness of tele-rehabilitation to in-person sessions.

As proved in other studies [27], comparing VHI with the GIRBAS scale (speech therapists' perceptual assessment), patients tend to increase their scores following therapy, but indicates that the therapy gives hope and increases psychological confidence to the patients.

3.3 Comparison between different software applications

In this section, the variations between three software applications used for vocal analysis are examined: the same acoustic parameters are extracted from identical voice recordings, using different software tools. This comparative analysis aims to evaluate the consistency of the results across platforms.

The software utilized in this thesis include Praat (the results of which were presented in the previous chapter), VOXPlot and MATLAB. Through a comparative analysis of these tools, the study aims to highlight any potential discrepancies or advantages that each software may offer in terms of usability and applicability in clinical or research settings, underlining the limitations of each.

3.3.1 Comparison between VOXPlot and Praat

The first comparison was conducted between VOXPlot and Praat, and the parameter extracted was the AVQI.

While for the analysis in Praat, Maryn's code [reported in Appendix A] was used to calculate AVQI, the automated analysis of VOXPlot eliminates the need of manual intervention. Users simply uploaded the desired audio files (in this case, 3 seconds of vowel and the reading passage) and selected the analysis language.

To analyze the differences between the two software tools, the delta for each parameter was calculated for all extracted data, combining both T0 and T1 measurements. For each parameter, the Δ was defined as the difference between the values obtained from Praat and those from Voxplot. The mean of these differences was determined, along with the corresponding standard deviation, to assess the variability between the two applications. A positive Δ indicates that, on average, the mean parameters obtained from Praat are higher than those from Voxplot, whereas a negative Δ signifies that the Praat values tend to be lower than those from Voxplot. Furthermore, the percentage deviation of each Voxplot value relative to Praat was computed using the absolute value of the difference between the parameters extracted through the two software, dividing by the corresponding Praat value; it allows for an effective valutation of the overall divergence between the two applications.

Table 3.10: VOXPlot AVQI results at T0

ID	CPPS$_{mean}$ (dB)	HNR (dB)	Shimmer (%)	Shimmer (dB)	Slope (dB)	Tilt (dB)	AVQI mean
1	9.40	11.44	14.34	1.30	-23.94	-7.44	6.30
3	6.81	11.59	14.63	1.39	-29.32	-3.80	8.65
5	12.29	13.80	10.11	0.97	-20.15	-8.06	4.29
6	9.73	10.52	12.87	1.22	-18.47	-9.39	5.73
8	9.43	10.66	14.90	1.35	-23.53	-9.10	5.95
11	6.16	9.86	16.18	1.47	-29.03	-8.86	7.73
13	8.76	11.69	13.94	1.32	-26.42	-6.72	6.86
15	9.46	12.59	10.56	1.02	-23.59	-5.00	6.58
19	11.99	11.31	9.66	1.00	-11.85	-7.37	5.03
20	8.49	12.17	12.84	1.22	-27.73	-7.64	6.54
2	10.95	13.34	11.92	1.16	-20.66	-6.83	5.61
4	10.41	11.88	13.69	1.25	-27.21	-8.52	5.34
7	9.54	10.18	13.91	1.29	-25.84	-8.28	5.98
9	10.70	12.17	11.98	1.14	-19.90	-8.27	5.34
10	11.41	12.69	11.26	1.05	-23.75	-8.31	4.68
12	12.36	12.82	11.21	1.09	-19.12	-9.55	4.08
14	3.96	7.03	18.56	1.61	-29.58	-7.64	9.35
16	11.11	13.46	11.13	1.03	-17.49	-8.32	4.96
17	9.93	10.21	10.58	0.93	-23.61	-6.08	5.83
21	10.53	14.27	9.35	0.90	-21.82	-7.29	5.23

Table 3.11: VOXPlot AVQI results at T1

ID	CPPS$_{mean}$ (dB)	HNR (dB)	Shimmer (%)	Shimmer (dB)	Slope (dB)	Tilt (dB)	AVQI mean
1	10.91	13.60	11.49	1.10	-25.58	-8.59	4.91
3	8.58	12.90	10.63	1.00	-25.46	-4.71	7.01
5	13.28	15.33	6.25	0.62	-17.62	-7.35	3.49
6	11.79	12.22	10.59	1.06	-16.57	-10.30	4.26
8	11.68	14.67	9.25	0.98	-15.08	-9.18	4.55
11	6.56	12.21	12.87	1.22	-27.33	-8.24	7.44
13	8.95	13.17	11.66	1.18	-26.36	-7.09	6.46
15	11.35	14.44	8.07	0.84	-17.84	-5.81	5.50
19	11.04	11.17	11.71	1.03	-17.70	-7.43	5.21
20	11.17	13.16	10.82	1.04	-25.24	-8.05	4.85
2	12.24	16.59	7.39	0.81	-11.45	-6.19	4.87
4	10.81	15.73	8.05	0.83	-26.85	-8.96	4.41
7	11.49	11.87	11.67	1.09	-24.88	-8.47	4.63
9	9.64	10.83	10.54	0.99	-18.95	-8.46	5.85
10	12.82	13.56	6.46	0.89	-12.23	-7.60	4.45
12	14.65	18.09	5.96	0.70	-11.08	-8.59	2.85
14	6.39	10.62	15.67	1.47	-25.51	-6.88	8.29
16	11.52	14.88	9.20	0.90	-15.49	-8.94	4.47
17	10.22	9.04	12.98	1.14	-22.21	-6.65	5.90
21	12.01	15.38	8.77	0.83	-20.43	-8.29	4.08

Table 3.12: Δ values between Praat and VOXPlot

	PRAAT VS VOXPLOT						
	CPPS (dB)	**HNR (dB)**	**Shimmer (%)**	**Shimmer (dB)**	**Slope**	**Tilt**	**AVQI**
Δ mean	-0.09	-0.2	-0.05	0.002	0.002	0.09	0.07
Δ std	0.11	0.67	0.86	0.07	0.28	0.14	0.14
dev %	1.31%	4.32%	5.83%	5.67%	0.82%	1.52%	2.56%

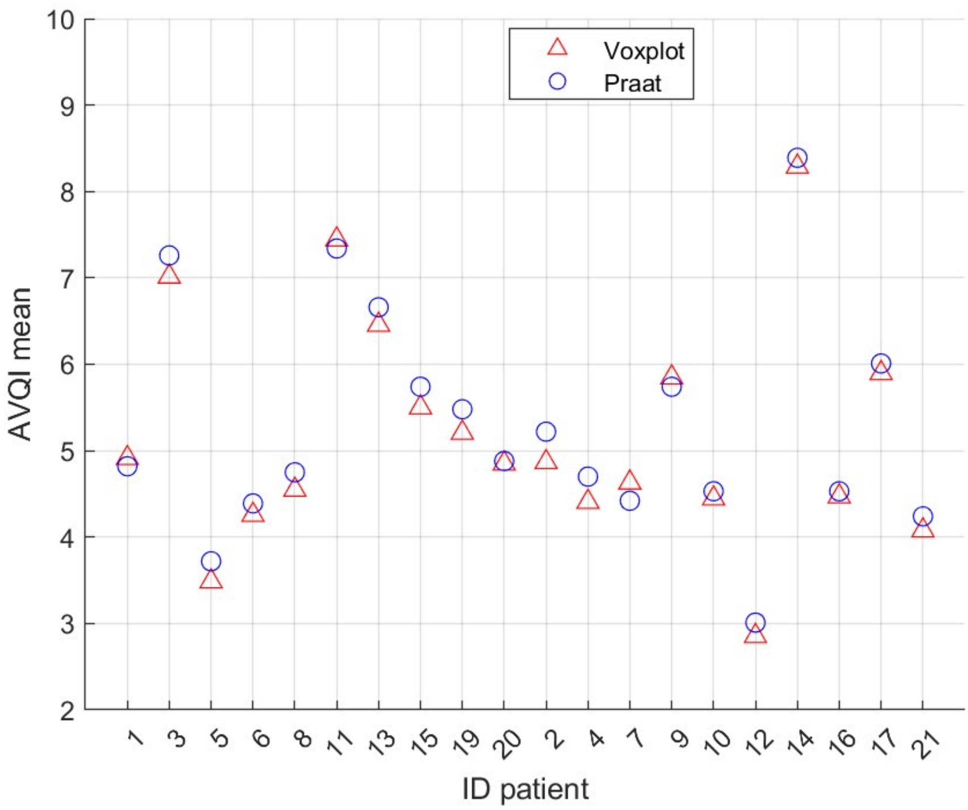

Figure 3.13: Comparison of AVQI extracted by Praat and Voxplot

As it can be observed in figure 3.13, AVQI values extracted from Praat and Voxplot are very similar:

- in Praat:
  AVQImean at T0 = 6.0 ± 0.29;
  AVQImean at T1 = 5.2 ± 0.29

- in VOXPlot:
  AVQImean at T0 =6.0 ± 0.30;
  AVQImean at T1 =5.1 ± 0.29.

These results align the expectations because, as written in the VOXPlot manual, the analysis in VOXPlot of the acoustic parameters and the two multidimensional indices (AVQI and ABI) is based on proven Praat algorithms with scientifically based presets.

The results confirm the reliability of both VOXPlot and Praat, demonstrating that they produce comparable outcomes in the analysis of vocal quality. While Praat is a highly versatile but complex software for acoustic analysis of a wide range of signals, requiring the use of specific algorithms, VOXPlot is specifically designed to the analysis of voice quality assessment. VOXPlot offers a user-friendly and efficient solution with an intuitive interface and providing a standardized and accessible option without compromising the accuracy of its analysis.

VOXPlot simplifies the process, but limits flexibility: user cannot modify input parameters as ranges or other settings, making it less customizable for advanced analysis.

In contrast, Praat allows to modify and personalize these input options, providing greater control over specific analytical parameters.

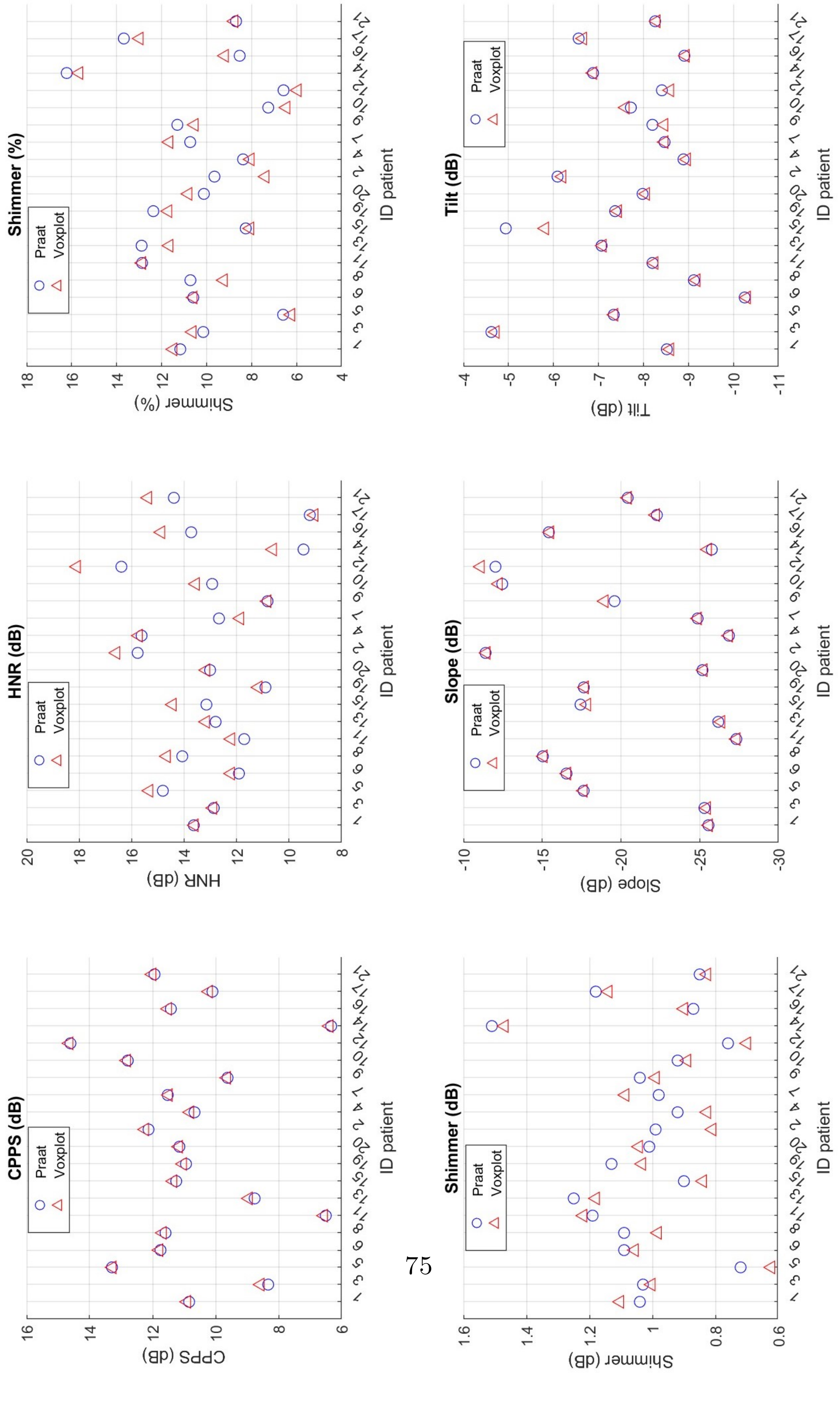

Figure 3.12: Comparison of parameters extracted by Praat and Voxplot

3.3.2 Comparison between Praat and Matlab

The following section presents a comparative analysis of the results obtained using Praat and Matlab.

In Praat, the extraction was performed using the script described in paragraph [], which allows automatic extraction of parameters through pre-implemented commands that must be configured with specific settings based on the analysis requirements. In Matlab, since it is not a software specifically designed for sound analysis, there are no pre-implemented commands to extract the acoustic parameters needed for voice analysis. For this reason, the development of a custom script (described in paragraph 2.7.1) and manual steps to preprocess the data and extract the parameters were required.

AVQI

The first investigated parameter is the AVQI, extracted from the same input signal: a concatenation of the sustained vowel and a reading passage of the phonetically balanced text "Notturno".

Also in this case, the index was calculated for three times for each patient, using three different repetitions of the sustained vowel /a/, and the average value was reported.

In tables 3.13 and 3.14 are shown the results obtained from the two software applications.

Table 3.13: MATLAB AVQI results at T0

ID	$CPPS_{mean}$ (dB)	HNR (dB)	Shimmer (%)	Shimmer (dB)	Slope (dB)	Tilt (dB)	AVQI mean
1	11.68	6.66	13.45	1.26	-24.85	-7.46	5.18
3	8.94	11.85	12.07	1.17	-29.17	-3.84	7.24
5	12.91	12.06	10.99	0.96	-20.13	-8.23	3.83
6	12.48	9.74	13.91	1.37	-18.41	-9.59	5.25
8	11.05	9.70	12.09	1.13	-23.37	-9.25	5.37
11	7.91	8.77	14.25	1.36	-28.85	-8.91	6.73
13	10.30	11.39	12.10	1.15	-26.28	-6.82	5.78
15	11.11	11.75	10.20	1.10	-23.45	-5.27	5.92
19	12.02	9.26	9.52	1.03	-11.60	-7.61	5.09
20	9.38	7.11	14.42	1.44	-27.70	-7.78	6.57
2	12.00	12.99	9.34	1.01	-20.54	-7.32	4.82
4	11.61	7.95	21.87	2.04	-26.42	-8.62	5.04
7	11.08	8.10	23.43	2.30	-25.72	-8.40	6.94
9	11.52	11.11	12.63	1.24	-19.89	-8.37	5.11
10	12.39	11.82	11.09	1.03	-23.73	-8.54	4.09
12	12.06	11.57	10.42	0.94	-19.13	-9.70	3.92
14	6.53	6.74	18.46	1.74	-29.48	-7.69	8.39
16	11.77	12.39	10.83	0.99	-17.39	-8.47	4.52
17	12.65	10.04	50.52	4.83	-23.59	-6.23	6.34
21	11.48	12.86	9.61	0.86	-21.81	-7.47	4.59

Table 3.14: MATLAB AVQI results at T1

ID	$CPPS_{mean}$ (dB)	HNR (dB)	Shimmer (%)	Shimmer (dB)	Slope (dB)	Tilt (dB)	AVQI mean
1	12.11	11.04	13.51	1.34	-25.54	-8.71	4.73
3	10.85	12.27	8.91	0.85	-25.30	-4.86	6.17
5	13.06	12.64	9.12	0.88	-17.50	-7.65	3.97
6	12.64	10.66	10.66	1.07	-16.54	-10.48	3.81
8	12.49	13.65	9.93	0.99	-14.64	-9.46	4.04
11	8.53	10.53	12.26	1.15	-27.30	-8.34	6.24
13	9.92	12.19	11.93	1.09	-26.16	-7.30	5.70
15	12.33	12.58	9.52	0.97	-17.43	-5.53	5.10
19	11.62	9.88	11.16	1.05	-17.63	-7.63	4.98
20	11.48	8.35	14.97	1.53	-25.20	-8.24	5.60
2	11.87	14.69	8.45	0.92	-11.44	-6.72	5.14
4	12.24	13.77	10.07	1.05	-26.66	-9.26	4.01
7	12.67	11.04	29.06	2.81	-24.87	-8.73	4.81
9	10.49	9.70	12.55	1.24	-19.60	-8.37	5.69
10	13.35	13.25	7.19	0.87	-12.47	-8.30	3.95
12	13.01	13.14	8.64	0.93	-12.07	-8.69	3.05
14	8.76	9.01	14.69	1.45	-25.77	-7.03	7.10
16	11.20	12.27	10.68	1.11	-15.46	-9.13	5.02
17	11.70	8.45	34.93	3.53	-22.29	-6.81	6.27
21	11.91	13.80	8.66	0.80	-20.44	-8.51	4.05

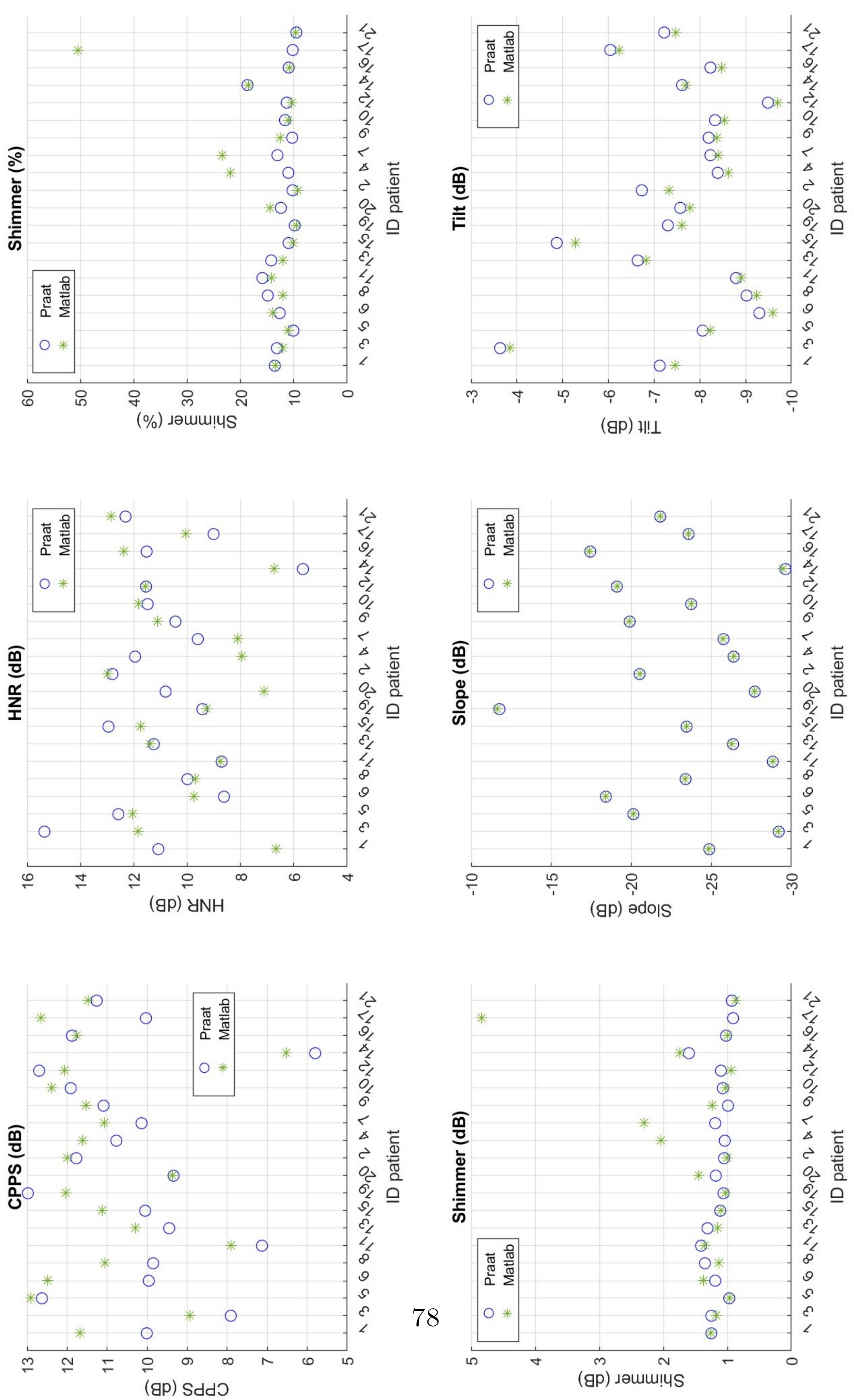

Figure 3.14: Comparison of parameters extracted at T0 by Praat and MATLAB

Figure 3.14 represents the comparison between the parameters extracted from MATLAB and Praat for patients at time T0.

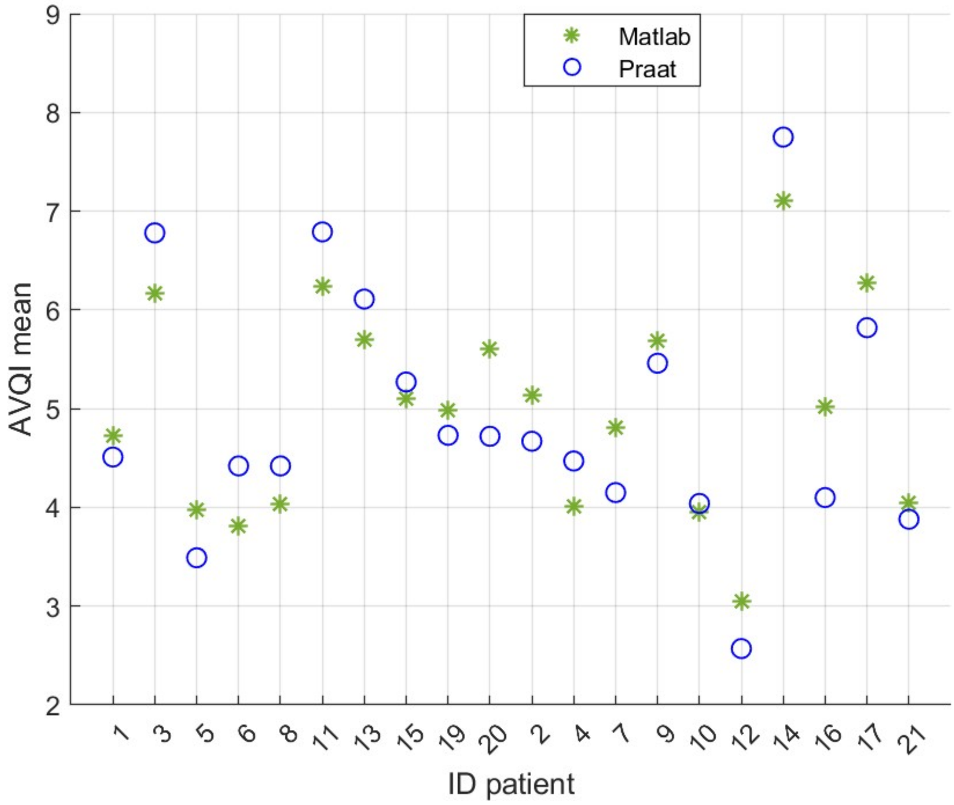

Figure 3.15: Comparison of AVQI extracted at T0 by MATLAB and Praat

Analyzing the data obtained from MATLAB and Praat from all the patients, ignoring the separation between the two rehabilitation groups, mean Δ value of each parameter has been evaluated, as difference as the value calculated through Praat and that extracted from MATLAB, the corresponding standard deviation, and the percentage deviation of each Matlab value relative to Praat:

Table 3.15: Δ values between Praat and MATLAB

	PRAAT VS MATLAB						
	CPPS (dB)	**HNR (dB)**	**Shimmer (%)**	**Shimmer (dB)**	**Slope**	**Tilt**	**AVQI**
Δ mean	- 0.39	0.45	- 2.66	- 0.28	- 0.03	0.27	0.03
Δ std	0.99	1.44	7.91	0.79	0.79	0.12	0.52
% dev	8.24%	8.49%	32.97%	35.55%	0.23%	3.80%	8.68%

The analysis of the table 3.15 reveals that all parameters produce comparable results between the two software tools, except for shimmer % and shimmer dB. In particular, figure 3.14 highlights a substantial difference for patient 17, where the outputs from the two software diverge significantly. This variation is probably due to an issue in Matlab, which seems to have incorrectly estimate the period of the vocal fold vibration.

After removing two outliers, the shimmer difference (both % and dB) was recalculated, resulting in more acceptable value:

- Shimmer %:
  - Δ mean = -2.66;
  - Δ std = 2.51;
  - % deviation = 16.18%.
- Shimmer dB:
  - Δ mean = -0.08;
  - Δ std = 0.28;
  - % deviation = 16.34%.

Although the shimmer values still show some discrepancy between the two software tools, the overall comparison can be considered acceptable.
This is supported by the fact that the difference between the AVQI values obtained is minimal, with the AVQI calculated using Matlab differing from that extracted using Praat by only 8.68%, with a Δmean of 0.03.

In summary, the differences between the two software can be shown as:

- in Praat:

  AVQImean at T0 = 5.6 ± 0.26;
  AVQImean at T1 = 4.9 ± 0.27

- in Matlab:

  AVQImean at T0 =5.5 ± 0.26;
  AVQImean at T1 =4.9 ± 0.23.

Sound Pressure Level (SPL)

The comparison between MATLAB and Praat includes also the calculation of the SPL value, extracted from the first 60 seconds of the free speech.
The two software, in addition to using different methods for silence removal, employ distinct approaches to calculate SPL:

- Matlab computes the RMS value for each 46 ms frame, which is then converted to SPL using the calibration formula. This results in an SPL vector from which various statistical measures, such as the mean, can be derived;
- Praat script calculates the RMS value of the entire vocal signal, producing a single value that is then converted to the average SPL of the signal using the calibration formula.

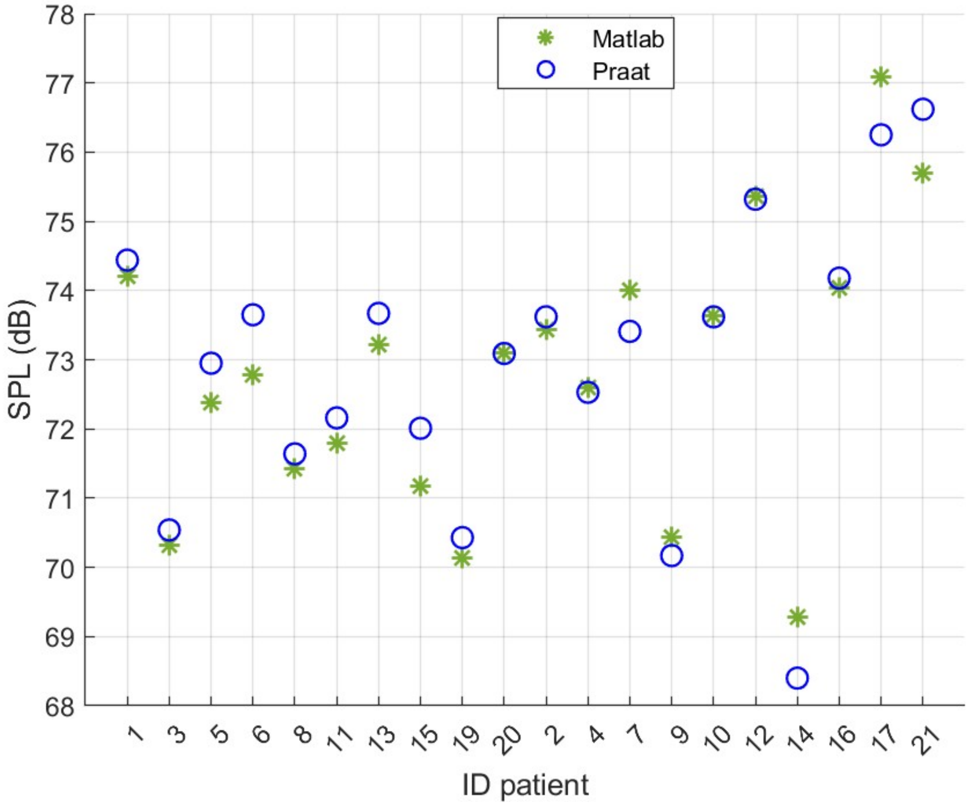

Figure 3.16: Comparison of SPL extracted at T0 by MATLAB and Praat

Figure 3.16 shows the SPL values extracted at T0 from Praat and for Matlab for all the patients.

To assess the agreement between the SPL values obtained from the two softare, the mean Δ is computed as the difference between Praat SPL and Matlab SPL: Δ mean found is 0.29 dB, indicating a very small average discrepancy between the two software. The Matlab average relative deviation from Praat of 0.87%, less than 1%, confirms the compatibility of the SPL measurements across both software tools.

Table 3.16: Comparison between **SPL** extracted in MATLAB and in Praat; *patients treated with tele-LSVT are reported in gray;*

	SPL @ 30 cm (dB)			
	T0		T1	
ID	**PRAAT**	**MATLAB**	**PRAAT**	**MATLAB**
1	70.93	71.22	74.44	74.21
3	69.16	69.21	70.54	70.33
5	72.21	72.08	72.95	72.38
6	73.91	73.08	73.65	72.78
8	68.26	68.25	71.64	71.42
11	75.51	72.99	72.16	71.79
13	70.38	69.92	73.67	73.22
15	70.38	70.13	72.01	71.17
19	68.36	66.86	70.43	70.13
20	72.29	66.98	73.09	73.10
2	72.54	71.83	73.62	73.43
4	76.41	75.72	72.53	72.60
7	70.26	70.64	73.41	74.00
9	70.67	70.80	70.17	70.44
10	73.23	73.16	73.62	73.63
12	71.48	71.42	75.32	75.37
14	66.6	67.61	68.4	69.28
16	70.5	70.15	74.18	74.03
17	74.62	76.75	76.25	77.09
21	73.44	73.22	76.62	75.69
mean	71.56	71.10	72.94	72.81
std. dev.	2.52	2.65	2.05	2.01

Chapter 4

Conclusions

In this thesis, different studies were conducted on vocal parameters and indexes extracted from recordings of patients with multiple sclerosis. The dataset was divided into two equal groups of ten subjects each: one group was treated with the standard LSVT-Loud therapy in clinic, while the other followed the same program through tele-rehabilitation (Tele-LSVT-Loud). For each subject, the speech material included three vocal tasks: three vocalizations of the sustained vowel /a/, reading of the phonetically balanced text "Notturno", and a minute of free speech. These tasks were simultaneously recorded using both an in-air microphone system and a contact microphone-based device (Vocal Holter, VH). Data were collected at two time points: T0, before the start of the rehabilitation program, and T1, following the completion of the treatment.

To evaluate the effectiveness of the LSVT-LOUD treatment, comparisons between T0 and T1 parameters and indexes were computed using Praat software; these analyses focused on AVQI, SPL, Warning Score and vowel parameters (jitter%, shimmer %, HNR and CPPS).

First, AVQI (Acoustic Voice Quality Index), a multiparametric index to assess voice quality, was calculated to verify the effectiveness of the treatment. Through the Praat script reported in Appendix A, AVQI was extracted from the concatenation of 3 seconds of sustained vowel and a phrase from the text reading. The change in AVQI (Δ AVQI), representing the difference between AVQI values at T0 and T1, showed a positive $\Delta AVQI$ of 0.74, suggesting a general improvement in voice quality among the patients. The SPL parameter, calculated on 60 s of free speech, also confirmed these results: SPL mean increased from 71.55 dB at T0 to 72.94 dB at T1 including all the patients. After excluding two patients considered outliers (whose ΔSPL were negative), the mean SPL increased from 71.07 dB to 73.00 dB, further confirming the effectiveness of the treatment. Δ parameters calculated on the sustained vowels followed a similar positive trend: jitter % and shimmer % showed average reductions at T1 (indicating healthier voices), and HNR and CPPS mean increased from T0.

When comparing the two rehabilitation modalities, it was observed that the two therapies can be considered equivalent in terms of the improvements in AVQI, SPL, and vowel parameters.

The Warning Score (WS), an index used to assess the state of vocal health status of patients, was also calculated using the vowel parameters extracted from the DAP unit of the Vocal Holter. Although the results of the WS between T0 and T1 were positive (indicating a reduction in the score, which corresponds to a less pathological voice), the values obtained were not consistent with those derived from the in-air microphone and processed with Matlab. Specifically, when comparing the parameters included in the WS, it was observed that while jitter, median CPPS and CPPS standard deviation values were comparable between the two microphones, shimmer values showed a discrepancy: on average, shimmer measured with the Vocal Holter was 44.5% higher than that measured from the in-air microphone. This discrepancy may be explained by the different source signals captured by the two microphones and their different characteristics (i.e. in terms of bandwidth), and suggests that new thresholds have to be established for the Vocal Holter. This contact microphone is in fact preferable, in term of insensitivity to other sound sources and patient comfort.

After the objective assessment of the effectiveness of the therapy, the VHI (Vocal Handicap Index) was used to quantify self-assessment of voice quality of patients and measure psycho-social effects of the vocal disorders during the treatment. Significant reductions in VHI are evident for many patients, indicating that patients, at the end of the rehabilitation program, feel generally better.

All these results bring us to the conclusion that both treatments improved MS patients' vocal status. Given the effectiveness of both methods of rehabilitation, telerehabilitation is preferable as it offer additional benefits for the patient. Telerehabilitation can be an alternative method to guarantee easier and continued access to healthcare services, minimising the barriers of distance, time and costs and allowing care to be provided directly at patient's home.

Further investigations were conducted in term of software applications, using the same recordings but processed with different tools: differences between Praat, VOXPlot and Matlab were examinated, highlighting positive aspects of each one.

Comparable outcomes were found from VOXPlot and Praat: the mean AVQI from VOXPlot differed only by 2.56% relative to the AVQI from Praat. These results are reliable since VOXPlot is based on Praat (version 6.0.48) algorithms. The results obtained through the script implemented in Matlab showed also minimal variation compared to Praat: AVQImean of MATLAB differed by 8.68% from that obtained through the Praat script.
Additionally, SPL was calculated using both software tools, obtaining very similar results (Matlab differs from Praat by 0.87%), supporting the consistency of these measurements across different platforms.

From the perspective of larger future projects, this study could be expanded with a larger patient dataset to obtain more robust and consistent results. Additionally, the dataset will be extended to include T2 follow-up recordings, which were not available at the time of this research, but will be essential for evaluating the long-term effects of the therapy. To further enhance the clinical relevance of the study, the GIRBAS scale (a perceptual assessment tool used by experts to evaluate the severity of dysphonia) could be incorporated alongside the VHI, providing valuable feedback from speech therapists. Moreover, a long-term monitoring study using the Vocal Holter will be conducted, enabling the collection of data over extended periods, even several hours, to capture more detailed vocal changes and monitor patients during daily activities. Finally, with the development of a new version of the Vocal Holter, new Matlab scripts could be integrated directly into the system. This enhancement could enable the device to extract more reliable parameters and automatically provide indices such as AVQI and Warning Score in real-time, improving its efficiency and usability in both clinical and tele-rehabilitation settings.

Appendix A

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.*"

Appendix B

AVQI Praat Script

ACOUSTIC VOICE QUALITY INDEX (AVQI) v.03.01

It is advocated to estimate someone's dysphonia severity in both continuous speech (i.e., 'cs') and sustained vowel (i.e., 'sv') (Maryn et al., 2010). This script therefore runs on these two types of recordings, and it is important to name these recordings 'cs' and 'sv', respectively.

This script automatically (a) searches, extracts and then concatenates the voiced segments of the continuous speech recording to a new sound; (b) concatenates the sustained vowel recording to the new sound, (c) determines the Smoothed Cepstral Peak Prominence, the Shimmer Local, the Shimmer Local dB, the LTAS-slope, the LTAS-tilt and the Harmonics-to-Noise Ratio of the concatenated sound signal, (d) calculates the AVQI-score based on the equation of Barsties & Maryn (2015). [...] To be reliable for the AVQI analysis, it is imperative that the sound recordings are made in an optimal data acquisition conditions.

Listing B.1: Praat Script for AVQI Calculation

```
# PART 0: HIGH-PASS FILTERING OF THE SOUND FILES
select Sound cs
Filter (stop Hann band)... 0 34 0.1
Rename... cs2
select Sound sv
Filter (stop Hann band)... 0 34 0.1
Rename... sv2

# PART 1: DETECTION, EXTRACTION AND CONCATENATION OF
```

```
# THE VOICED SEGMENTS IN THE RECORDING OF CONTINUOUS SPEECH
select Sound cs2
Copy... original
samplingRate = Get sampling frequency
intermediateSamples = Get sampling period
Create Sound... onlyVoice 0 0.001 'samplingRate' 0
select Sound original
To TextGrid (silences)... 50 0.003 -25 0.1 0.1 silence sounding
select Sound original
plus TextGrid original
Extract intervals where... 1 no "does not contain" silence
Concatenate
select Sound chain
Rename... onlyLoud
globalPower = Get power in air
select TextGrid original
Remove

# Loop through voiced segments and concatenate
select Sound onlyLoud
signalEnd = Get end time
windowBorderLeft = Get start time
windowWidth = 0.03
windowBorderRight = windowBorderLeft + windowWidth
globalPower = Get power in air
voicelessThreshold = globalPower*(30/100)

while windowBorderRight < signalEnd - windowWidth:
    Extract part... 'windowBorderLeft' 'windowBorderRight' Rectangular
   1.0 no
    partialPower = Get power in air
    if partialPower > voicelessThreshold:
        call checkZeros(0)
        if zeroCrossingRate != undefined and zeroCrossingRate < 3000:
            Concatenate
            Rename... onlyVoice
    windowBorderLeft += 0.03
    windowBorderRight += 0.03

# PART 2: DETERMINATION OF THE SIX ACOUSTIC MEASURES AND AVQI
   CALCULATION
```

```
# Slope of the long-term average spectrum
select Sound avqi
To Ltas... 1
slope = Get slope... 0 1000 1000 10000 energy

# Tilt of trendline through the LTAS
select Ltas avqi
Compute trend line... 1 10000
tilt = Get slope... 0 1000 1000 10000 energy

# Amplitude perturbation measures (Shimmer)
select Sound avqi
To PointProcess (periodic, cc)... 50 400
percentShimmer = Get shimmer (local)... 0 0 0.0001 0.02 1.3 1.6
shdb = Get shimmer (local_dB)... 0 0 0.0001 0.02 1.3 1.6

# Harmonic-to-noise ratio (HNR)
select Sound avqi
To Pitch (cc)... 0 75 15 no 0.03 0.45 0.01 0.35 0.14 600
hnr = Voice report... 0 0 75 600 1.3 1.6 0.03 0.45

# Final AVQI Calculation
avqi = (4.152-(0.177*cpps)-(0.006*hnr)-(0.037*percentShimmer)+(0.941*
    shdb)+(0.01*slope)+(0.093*tilt))*2.8902
```

Bibliography

[1] National MS Society. *MS Prevalence*. Accessed: 2024-09-05. URL: `https://www.nationalmssociety.org/about-the-society/who-we-are/research-we-fund/ms-prevalence` (cit. on p. 1).

[2] Shepherd Center. *What is MS? - Types of Multiple Sclerosis*. Accessed: 2024-07-09. 2024. URL: `https://www.shepherd.org/programs/outpatient/multiple-sclerosis/what-is-ms/types` (cit. on p. 2).

[3] 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 (Jan. 2023), pp. 1–11 (cit. on pp. 3, 7, 25).

[4] Svensson P. Hartelius L. «Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: a survey». In: *Folia Phoniatrica et Logopaedica* 46 (Nov. 1994), pp. 9–17 (cit. on p. 3).

[5] Zorica Aleric, Ervin Jancic, and Vladimir Bauer. «Comparing Voice Self-Assessment with Auditory Perceptual Analysis in Patients with Multiple Sclerosis». In: *International Archives of Otorhinolaryngology* 19.02 (Dec. 2014), pp. 100–105 (cit. on p. 3).

[6] National Multiple Sclerosis Society. *Speech Problems in MS*. Accessed: 2024-09-05. 2024. URL: `https://www.nationalmssociety.org/understanding-ms/what-is-ms/ms-symptoms/speech-problems#section-2` (cit. on p. 3).

[7] Zhaoyan Zhang. «Mechanics of human voice production and control». In: *The Journal of the Acoustical Society of America* 140.4 (Oct. 2016), pp. 2614–2635 (cit. on p. 4).

[8] C. Hrelec and E. Zhang. «Anatomy and Physiology of Phonation». In: *International Journal of Head and Neck Surgery* 12.4 (2021), pp. 125–130 (cit. on p. 4).

[9] Paul Warren and Phil Rose. *VOXplot Manual*. Accessed: 2024-08-20. 2019. URL: https://voxplot.lingphon.com/downloads/VOXplot_Manual.pdf (cit. on pp. 6, 22).

[10] C. Baldanzi, V. Crispiatico, S. Foresti, E. Groppo, M. Rovaris, D. Cattaneo, and C. Vitali. «Effects of Intensive Voice Treatment (The Lee Silverman Voice Treatment [LSVT LOUD]) in Subjects With Multiple Sclerosis: A Pilot Study». In: *Journal of Voice* 36.4 (2022), 585.e1–585.e13 (cit. on p. 6).

[11] Linda A. Bryans, Andrew D. Palmer, Shannon Anderson, Joshua Schindler, and Donna J. Graville. «The impact of Lee Silverman Voice Treatment (LSVT LOUD®) on voice, communication, and participation: Findings from a prospective, longitudinal study». In: *Journal of Communication Disorders* 89 (Jan. 2021), p. 106031 (cit. on p. 6).

[12] F. Rossetto, F. Borgnis, S. Isernia, E. Foglia, E. Garagiola, O. Realdon, and F. Baglio. «System Integrated Digital Empowering and teleRehabilitation to promote patient Activation and well-Being in chronic disabilities: A usability and acceptability study». In: *Frontiers in Public Health* 11 (Mar. 2023) (cit. on p. 7).

[13] Fondazione Don Carlo Gnocchi. *La Storia - Il Nostro Percorso*. Accessed: 2024-08-02. URL: https://www.dongnocchi.it/la-fondazione/la-storia/il-nostro-percorso (cit. on p. 8).

[14] 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 p. 10).

[15] James Hillenbrand, Ronald Cleveland, and Robert Erickson. «Acoustic Correlates of Breathy Vocal Quality». In: *Journal of speech and hearing research* 37 (Sept. 1994), pp. 769–78 (cit. on pp. 11, 12, 36).

[16] James Hillenbrand and Robert A. Houde. «Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech». In: *Journal of Speech, Language, and Hearing Research* 39 (Apr. 1996), pp. 311–321 (cit. on p. 12).

[17] 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 (2018), pp. 646–654 (cit. on p. 12).

[18] Jan G. Švec and Svante Granqvist. «Tutorial and Guidelines on Measurement of Sound Pressure Level in Voice and Speech». In: *Journal of Speech, Language, and Hearing Research* 61.3 (Mar. 2018), pp. 441–461 (cit. on pp. 13, 14).

[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. 17).

[20] M. Fantini and A. et al. Ricci Maccarini. «Validation of the Acoustic Voice Quality Index (AVQI) Version 03.01 in Italian». In: *Journal of Voice* 37.4 (July 2023), 631.e1–631.e6 (cit. on p. 18).

[21] Arianna Astolfi, Alessio Carullo, Giuseppina Puglisi, V.I. Fissore, and J. Clerici. «An Experimental Approach for the Evaluation of Vocal Behavior of University Professors». In: *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*. Jan. 2022, pp. 449–456 (cit. on pp. 20, 61).

[22] Paul Boersma and David Weenink. *Praat: Doing Phonetics by Computer*. Accessed: 2024-07-25. University of Amsterdam. 2024. URL: `https://www.fon.hum.uva.nl/praat/` (cit. on p. 21).

[23] Paul Boersma and Vincent Van Heuven. «Speak and unSpeak with PRAAT». In: *Glot Int* 5 (Jan. 2001), pp. 341–347 (cit. on p. 21).

[24] Ben Barsties and Youri Maryn. «External Validation of the Acoustic Voice Quality Index Version 03.01 With Extended Representativity». In: *Annals of Otology, Rhinology & amp; Laryngology* 125 (Mar. 2016), pp. 571–583 (cit. on pp. 38, 42).

[25] 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 (Mar. 2018), pp. 1591–1602 (cit. on p. 43).

[26] Barbara H. Jacobson, Alex Johnson, Cynthia Grywalski, Alice Silbergleit, Gary Jacobson, Michael S. Benninger, and Craig W. Newman. «The Voice Handicap Index (VHI): Development and Validation». In: *American Journal of Speech-Language Pathology* 6.3 (Aug. 1997), pp. 66–70 (cit. on p. 45).

[27] Sara Palmieri. «Assessment of the vocal status of multiple sclerosis patients: comparison with healthy subjects and evaluation of vocal rehabilitation». MsC thesis. Politecnico di Torino, 2023 (cit. on p. 70).