POLITECNICO DI TORINO

Master's Degree in Biomedical Engineering, Biomedical Instrumentation

Master's Degree Thesis

In-Clinic and Tele-Rehabilitation Paths for Multiple Sclerosis Patients with Speech Disorders

Comparison through vocal signals acquired by in-air and contact microphones

Supervisors

Prof. Alessio CARULLO

Prof. Alberto VALLAN

Candidate

Giulia SCHIAZZA

OCTOBER 2024

Abstract

Voice is essential for communication, relationship building, and emotional expression: its production involves the vocal cords, respiratory system, and upper vocal tract, and improper control can lead to dysphonia. Multiple Sclerosis (MS) is an immune-mediated demyelinating disease of the central nervous system that can result in physical disabilities, including voice disorders, in fact 62% of MS patients experience hypophonia. The Don Gnocchi Foundation supports patient's rehabilitation, mainly with the Lee Silverman Voice Treatment (LSVT)-Loud, an effective therapy based on motor learning and neural plasticity principles. Traditionally offered in person, LSVT-Loud has recently transitioned online due to COVID-19, allowing patients to receive therapy at home without needing transportation or accompaniment, which improves accessibility and convenience. Therefore, given the advantages of online rehabilitation for patients, it has been decided to focus this study on verifying the non-inferiority of tele-rehabilitation compared to in-person one. To address this objective, the Don Carlo Gnocchi Foundation provided a dataset of 20 MS patients, 10 treated with tele-rehabilitation and 10 with in-clinic rehabilitation. Each patient is asked to produce three repetitions of the sustained vowel /a/, a monologue, and a reading of the poem 'Notturno' at three different times T0 (pre), T1(post), and T2 (follow-up). This vocal material is recorded using two different devices, a vocal recorder equipped with an in-air microphone and the Vocal Holter (VH) device, which include a contact microphone. In addition to the objective of assessing the non-inferiority of tele-rehabilitation, this study also aims to compare the two different recording systems. The recordings of in-air microphone are first cleaned removing artefacts using Audacity, then they are pre-processed in MATLAB and eventually the vocal features are extracted. The main parameters extracted for comparison between VH and in-air microphone, or tele-rehabilitation and in-clinc rehabilitation, include fundamental frequency, jitter, shimmer, cepstral peak prominence smoothed (CPPS), harmonic-to-noise ratio (HNR), and sound pressure level (SPL). In this study, unlike previous work, the pre-processing phase uses varying thresholds to differentiate between harmonic and non-harmonic frames. While a 0 dB threshold is previously used for healthy subjects, thresholds of -10 dB and -20 dB are investigated for MS patients to account for disease-specific non-harmonic frames. To demonstrate the results evaluated with different thresholds, the mean value of HNR at T0 is computed, obtaining for a 0 dB threshold the value of 7.8 dB. Compared to this value, using the -10 dB and -20 dB thresholds results in HNR differences of -1.8 dB and -1.9 dB respectively. In the case of the VH, it generates a text file containing all the parameters extracted during the recording. By analyzing parameters trends across different time it is possible to

assess whether rehabilitation has led to improvement and to determine which of the two techniques is more effective or if they have comparable effects. Among the analyzed parameters, CPPS shows the most similar mean values between the presence and tele-rehabilitation conditions: the mean difference between T1 and T0 for rehabilitation in presence is 0.33 dB, while in the online case it is 0.35 dB. As a result, it has been demonstrated that tele-rehabilitation and in-clinc rehabilitation are comparable, showing similar benefits.

Acknowledgements

A Mamma e Papà, il dono più grande che mi è stato fatto dalla vita.
Non sarei mai arrivata fin qui se non fosse stato per voi, vi amo.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

What does it mean to communicate? It comes from the Latin word 'communicare', implying sharing with others.

> "Aware of his own responsibility and strong in its role, communication is a social expression, a putting of value at the service of someone or something outside of itself: it is not enough to pronounce, write or draw to communicate; communication occurs when it arrives, when expression is understood and becomes common heritage for the construction of a discussion, of knowledge, of a culture."[1]

Communication is fundamental, as it allows people to express their feelings and facilitates connections between individuals. However, in some conditions, such as Multiple Sclerosis, people may struggle with this ability due to issues that also affect the phonation system.

1.1 Phonation System

How people produce voice? It's possibile thanks to three different systems in the human body:

- **Air Pressure System:** It generates and regulates the flow of air and is composed of trachea, diaphragm, chest muscles, abdominal muscles, ribs and lungs.
- **Phonatory System:** The mechanical and aerodynamic energy coming from the previous system is converted into acoustic energy thanks to the interaction between air and moving walls of the larynx.

- **Articulatory System:** At this point, the sound waves produced travel in the vocal tract where the cavities filter the acoustic signal until there is production of the voice through the mouth and/or nose. The vocal tract is composed of two spaces: the buccal and nasal cavities. The first one includes the space between the glottis and the lips, and can be altered by the movement of tongue, jaw and lips, known as the speech articulators. These are responsible for the alteration of speech by modifying this space. Otherwise, the nasal cavity does not change its shape, so it is only the soft palate that controls the passage or not of sounds through the nose.

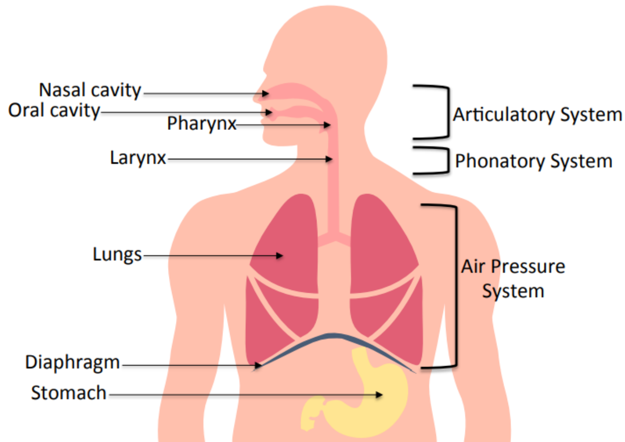


Figure 1.1: Voice production apparatus

Summarizing: the acoustic waves are produced by the interaction between the air pressure system and the phonatory system and then these waves became voice sounds thanks to the articulatory system. [2]

The ability to produce purposeful vocalizations and speak fluently is regulated by a complex network of mechanisms originating in the central nervous system (CNS). Central regulation plays a crucial role in voice production and speech, involving a range of specialized mechanisms that coordinate advanced cortical processing, brainstem reflexes, and peripheral nerves. Voice production is primarily controlled by the speech motor cortex, which includes both the laryngeal motor cortex (LMC) and the orofacial motor cortex. This cortical area manages over 100 muscles involved in phonation, swallowing, and breathing. Neuroimaging and electrical stimulation techniques are used to localize the LMC, which is situated in area 4 of the motor cortex in humans. This differs from nonhuman primates, where the laryngeal motor area is located in area 6 of the premotor cortex.

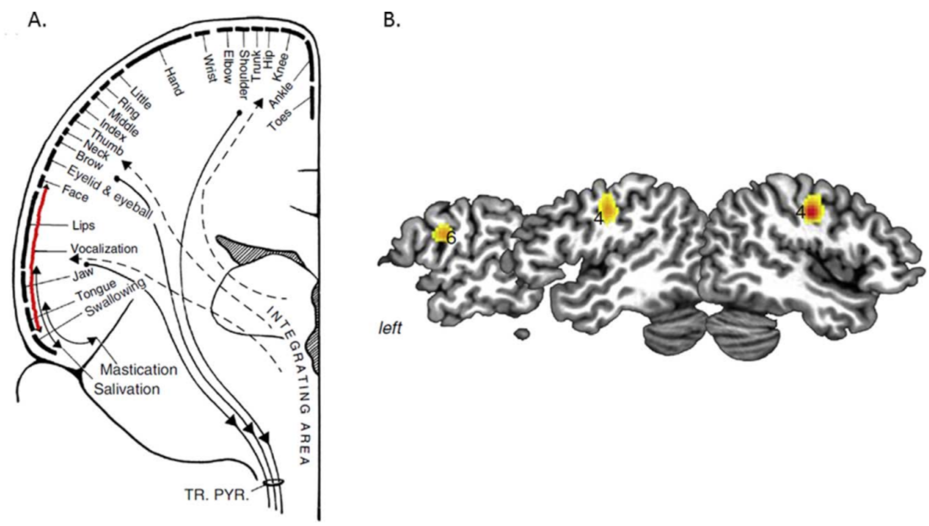


Figure 1.2: A. Motor sequence within the primary motor cortex, with the vocalization region located in the lower part of the precentral gyrus. B. Functional magnetic resonance imaging during voice production showed activation in the laryngeal motor cortex, with prominent activation in area 4 and less in area 6 [3]

In humans, the LMC in area 4 allows for direct connections between the LMC and the laryngeal motoneurons in the nucleus ambiguus of the brainstem. This direct connectivity facilitates faster and more precise coordination of complex laryngeal, orofacial, and respiratory movements, likely enhancing the learning and voluntary control of vocalization for speech purposes. [3]

1.2 Multiple Sclerosis

Multiple sclerosis (MS) is a chronic autoimmune inflammatory and demyelinating disease of the central nervous system. Immune system cells attack the myelin in the central nervous system, leading to the formation of plaques visible through magnetic resonance imaging (MRI). MS primarily affecting individuals between the ages of 20 and 40, with an average age of onset around 30 years. The incidence between women and men ranges from 2:1 to 3:1 and continues to increase [4]. The main reasons for this disparity include:

- **Hormones:** Female susceptibility to MS may be influenced by both the detrimental effects of female sex hormones and the protective effects of male sex hormones.
- **Genetic factors:** Women have XX chromosomes and men have XY chromosomes. Some genes on the Y chromosome may reduce susceptibility to MS,

while some genes on the X chromosome may promote the disease. Additionally, X chromosome genes in males come only from the mother, whereas in females they can come from both parents, which may help explain the higher incidence in women.

- **Epigenetics and gene-environment interaction:** Environmental factors and epigenetic mechanisms can influence gene expression and contribute to the differences in incidence between the sexes.

Despite the higher incidence and more robust immune responses in women, the disease course in this sex is not more aggressive nor is the prognosis poorer compared to men. If the central nervous system (CNS) in women is indeed more resistant to neurodegeneration than in men, then understanding the mechanisms behind this sex-related difference could lead to the development of neuroprotective therapies.

MS is usually suspected after the first neurological disturbance, known as clinically isolated syndrome (CIS), approximately 30% of CIS cases may experience a relapse within a year and potentially progress to clinically definite multiple sclerosis (CDMS). Diagnosing MS is complex, as there is no single test, and it relies on various criteria due to the high variability between patients in terms of both symptoms and disease progression. In the past, the diagnosis was based on clinical symptoms, neurophysiological exams, and cerebrospinal fluid (CSF) analysis, searching for signs of involvement in multiple areas of the central nervous system (spatial dissemination, DIS) and lesions occurring at different times (temporal dissemination, DIT). However, this diagnostic process could take years.

The discovery of drugs capable of slowing the progression of the disease necessitated a faster diagnosis. The introduction of magnetic resonance imaging (MRI) enhanced the ability to detect the typical MS lesions, integrating clinical data. In 2001, the McDonald criteria are introduced, defining MS diagnosis based on the essential characteristics of the disease: DIS and DIT. These criteria are later revised in 2005, 2010, and 2017 [5].

The McDonald criteria focus on:

- **Clinical interpretation:** The interpretation and integration of clinical history, physical examination, and imaging and laboratory test results by a clinician with expertise in multiple sclerosis are fundamental for a reliable diagnosis.

- **MRI:** The use of magnetic resonance imaging (MRI) has increased to support the diagnosis of multiple sclerosis and to identify atypical radiological features that could argue against this diagnosis. Standardized MRI protocols have been established for the diagnostic process, to determine prognosis, and for follow-up. Brain and spinal cord MRI remains the most useful paraclinical test to assist in the diagnosis of multiple sclerosis, potentially substituting for clinical findings in determining DIS or DIT in patients with a typical clinically isolated syndrome. Juxtacortical and cortical lesions are among the main indicators of the presence of MS, but it is crucial to be cautious, as cortical lesions can be easily confused with neuroimaging artifacts.

- **CSF:** In adult patients presenting with a clinically isolated syndrome, the presence of oligoclonal bands in the cerebrospinal fluid (CSF) proteins produced by immune system cells can indicate a high risk of a second attack, thereby supporting the diagnosis.

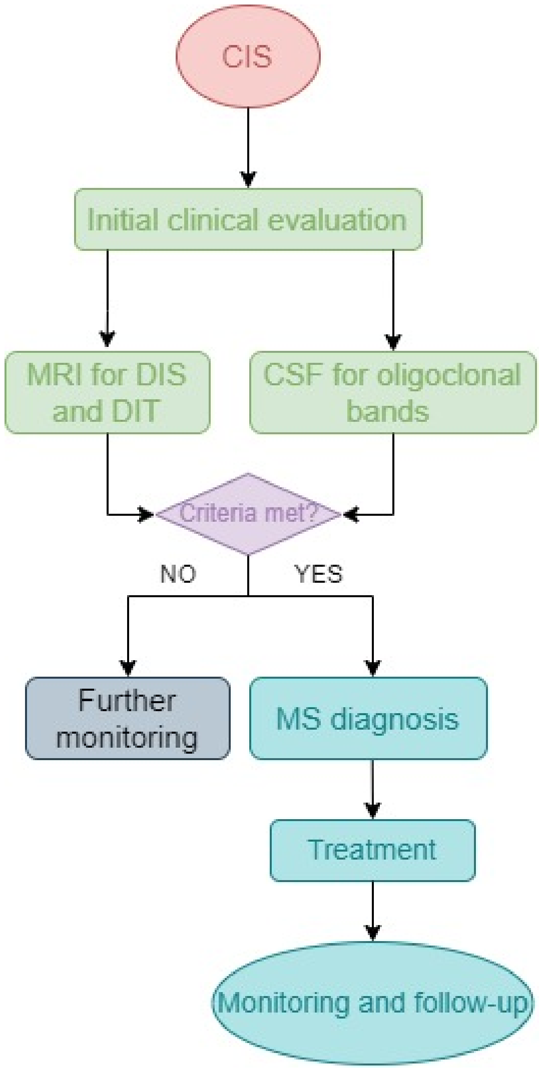
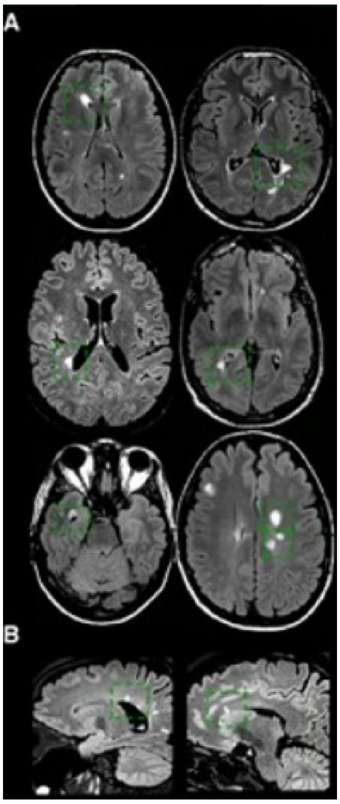
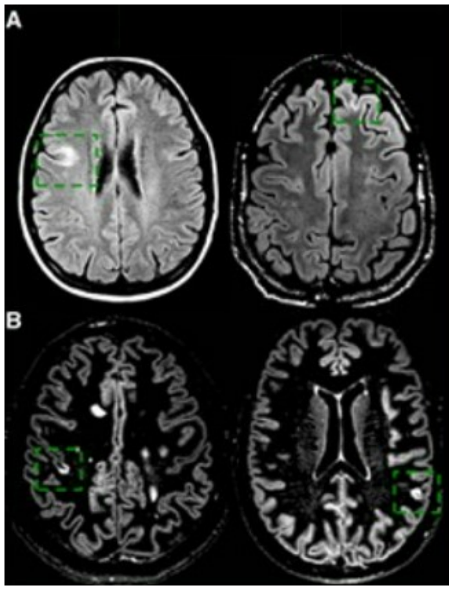


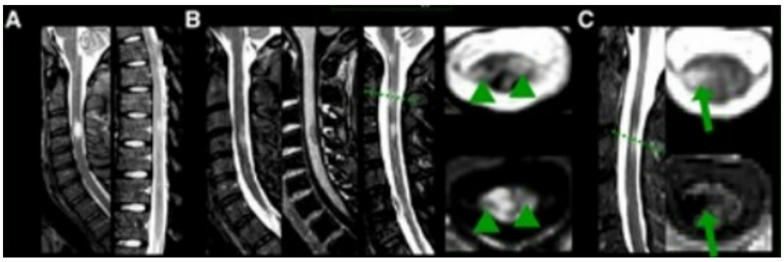
Figure 1.3: Flowchart illustrating the diagnostic and treatment process for multiple sclerosis.



(1)



(2)



(3)

Figure 1.4: Characteristics of typical multiple sclerosis lesions: (1) Periventricular lesions (A) examples of periventricular lesions suggestive of multiple sclerosis; (B) periventricular lesions perpendicular to the corpus callosum; (2) Cortical/juxtacortical lesions examples of (A) juxtacortical lesions and (B) cortical lesions suggestive of multiple sclerosis; (3) Spinal cord lesions, (A) Lesions visible in the cervical and thoracic cord on MRI, (B) Cervical cord lesions showing dark areas on imaging, (C) A cervical cord lesion affecting the lateral column and central grey matter [6].

The symptoms of MS vary depending on the severity of the inflammatory response and the location and extent of the plaques, which primarily appear in the brainstem, cerebellum, spinal cord, optic nerves, and white matter around the cerebrospinal fluid-filled ventricles. As a direct consequence, patients affected by MS may experience motor and sensory disturbances, visual disabilities, cognitive impairments, and speech and language deficits.

The effect of the disease on one's central system becomes quite evident by analyzing their phonatory capabilities. Specifically, the disease affects the subsystems of speech, including respiration, phonation, articulation, and prosody. The lack of voluntary coordination of muscle movements contributes to these problems, resulting in noticeable difficulties in speech fluency and speed.

Although there is currently no cure for MS, treatment can alleviate symptoms, prevent further relapses, and improve the quality of life for patients. [7]

This work is focused on the evaluation of vocal problems for MS patients and the comparison between tele-rehabilitation and rehabilitation in presence for these subjects.

1.3 The Importance of Tele-rehabilitation

Diseases, such as MS, require rehabilitation, which is essential for improving the quality of life for patients. These services are often not exploited, despite their importance, due to various problems such as services or resources of patients are scarce, the high demand leads to saturation of services and generation of waiting lists and many patients need someone to accompany them in the facilities to be able to use these services. So, especially during the SARS-CoV-2 infection, it is crucial to implement tele-rehabilitation techniques to ensure patients receive the care they need. [8] Therefore, as a consequence of these problems, in collaboration with the Don Carlo Gnocchi foundation, voice tele-rehabilitation techniques have been developed for patients with MS and the objective of this work is to assess the non-inferiority of tele-rehabilitation compared with rehabilitation in presence.

Don Carlo Gnocchi foundation is one of the most prominent foundations in the field of medical research, which has always focused on all forms of disability, with particular attention to rehabilitation techniques. The development of this project began to take shape in 1945, after the dramatic experience of World War II. Don Carlo Gnocchi had the desire to "start from the least fortunate, to redeem their innocent suffering and build a hope for the future." Thanks to Don Carlo Gnocchi's mission, there are now cutting-edge facilities that assist many sick people.[9]

For this thesis work, collaboration with the foundation is essential, as they provide audio files of MS patients to evaluate two different rehabilitation techniques and assess their effectiveness, while also supplying voice recordings made with two different types of microphones: a contact microphone and an in-air microphone, allowing for comparison between them.

Chapter 2

Materials and Methods

In this work all patients are treated with LSVT-Loud (Lee Silverman Voice Treatment) which is different from the conventional speech therapy techniques. LSVT-Loud initially is used for hypophonia in Parkinson's disease (PD) but Crispiatico et al. [10] have also demonstrated the effectiveness of the treatment for patients with MS. LSVT-LOUD focuses on increasing voice through intensive and repetitive exercises, improving muscle activation to promote greater vocal amplitude in everyday life, it is based on principles that stimulate neuronal plasticity and brain reorganization. The rehabilitation includes sessions with daily tasks and hierarchical exercises: daily tasks involve 30 minutes of sustained /a/ phonation, high-volume /a/, pitch glides, and the reading of 10 functional sentences; hierarchical exercises consist of 30 minutes of reading and conversation, gradually increasing in difficulty through longer durations and more complex task. After demonstrating the effectiveness of this treatment for MS patients, efforts are made to make it more accessible and to simplify the lives of people with MS, thanks to telerehabilitation in order to offer a care service in a comfortable way at home, incorporating rehabilitation into the individual's daily activities and social interactions, while encouraging self-care management. Vitali et al. [11] provide a protocol to compare Tele and in-clinc LSVT-Loud on which this study is focused, the only difference between telerehabilitation and in-clinic rehabilitation lies in the delivery method of the rehabilitation treatment. In the first case, a digital platform (Maia Platform) is used, accessible to both the speech therapist and the patient, while in the second case, the patient must go to the clinic and perform the necessary activities with the therapist.

2.1 Data set

Participants will be recruited from the MS Rehabilitation Unit of the IRCCS Don Carlo Gnocchi Foundation ONLUS of Milan (Italy). The selection of patients for this study is be based on specific inclusion and exclusion criteria [11].

Inclusion Criteria

- Age over 18
- Diagnosed with Multiple Sclerosis (MS) based on McDonald's criteria
- Mild to severe voice symptoms confirmed by two speech-language therapists
- Mini-Mental State Examination (MMSE) score greater than 24 [12], a test to assess cognitive functions, such as memory and orientation.
- Access to a personal computer and internet at home
- Stable medication treatment with dopamine agonists and/or steroids for the last 3-6 months (if applicable)
- Signed informed consent to participate in the study

Exlusion Criteria

- Dysphonia related to other diseases
- Presence of other neurological conditions (besides MS)
- History of laryngeal cancer, radiotherapy, head/neck trauma, or intubation
- Visual or hearing problems
- Major psychiatric comorbidities
- Participation in voice rehabilitation sessions (conventional treatment or LSVT-Loud) in the last 6 months

After verifying the inclusion criteria, the data set consists of 20 MS patients, who are randomly assigned to either the experimental or control group, with the experimental group receiving Tele LSVT-Loud therapy and the control group undergoing traditional in-clinic LSVT-Loud therapy.

Patient ID	Tele-rehabilitation	In-clinic rehabilitation
1	-	X
2	X	-
3	-	X
4	X	-
5	-	X
6	-	X
7	X	-
8	-	X
9	X	-
10	X	-
11	-	X
12	X	-
13	-	X
14	X	-
15	-	X
16	X	-
17	X	-
18	X	-
19	-	X
20	-	X
21	X	-

Table 2.1: List of patients assigned to tele-rehabilitation and in-clinic rehabilitation.

In the Table, it is noted that there are 21 patients, not 20, because patient 21 is included as a backup in case there are data issues with other patients, specifically patient 18 at T1, which is excluded from the evaluation in this instance.

Therefore, after establishing the data set and dividing the patients between in-clinc and tele-rehabilitation, the study can begin. To obtain results and achieve the objective of this work, patients are asked to produce three different vocal materials:

- **three repetitions of the sustained vowel /a/**
- **a monologue**
- **the reading of "Notturno"**

In order to assess the non-inferiority of tele-rehabilitation compared to in-clinic rehabilitation and to compare the different microphones used, these vocal materials are evaluated at three different times:

- **T0**: baseline
- **T1**: after treatment
- **T2**: follow-up 3 months after T1

So all the trial work plan is summarized in the following Figure, which provides a comprehensive overview of the study's methodology, including the recruitment process, treatment timelines, and assessment points, allowing for a clearer understanding of the overall research framework.

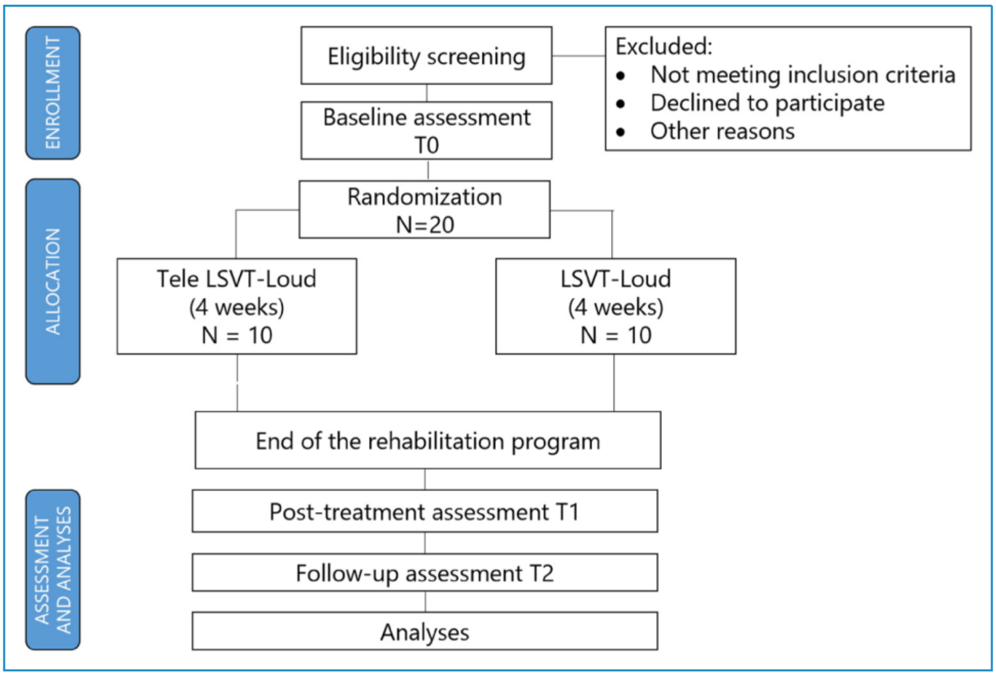


Figure 2.1: The trial work plan [11]

2.2 Equipment and setup

All voice signals are acquired using both a vocal recorder with an embedded microphone in air and the Vocal Holter (VH) device, which includes a contact microphone, in a quiet room to minimize ambient noise. The Roland R-05 is a portable system that includes the in-air microphone, positioned 30 cm from the mouth, it features a 16-bit resolution and a sampling rate set to 44.1 kSa/s, it produces a .wav file, which is analyzed using MATLAB scripts to extract all the required parameters.

Figure 2.2: Roland R-05 portable audio recorder used for in-air voice signal acquisition.

Differently, the VH measures the activity of the vocal cords through the skin vibrations [13]. It's made up of these elements:

- **DAP unit which embeds an audio microphone and spacer**
- **Contact microphone**
- **Power adapter and cable**
- **Instruction manual**

Figure 2.3: Vocal Holter Med kit[13]

Similar to the in-air microphone, the sampling rate is 44.1 kSa/s and the resolution is 16-bit. The samples acquired with the VH device are grouped into frames of 46 ms, and only voiced frames are evaluated. It is crucial to position the device around the subject's neck, ensuring that the ends of the collar make as much contact as possible with the skin area above the vocal cords. Additionally, the device should be comfortable for the user, as it must remain in place throughout the entire recording period. Since each subject has different body characteristics, the VH device is equipped with an adjustable pin to either loosen or tighten the collar based on neck size. This allows for both long and short measurements to evaluate fatigue, making it essential for the device to be comfortable, as the patient may need to wear it for several hours. The device produces a text file containing all the processed parameters.

Figure 2.4: Correct positioning of the vocal Holter.[13]

2.3 Parameters

To perform an objective analysis of dysphonia, various parameters must be extracted. These parameters can be grouped into three domains:

- **Time**
- **Frequency**
- **Cepstral**

In the time domain, jitter and shimmer are commonly used and frequently cited in the literature. The issue with these parameters is that they depend heavily on the accurate identification of vocal fold vibration cycles; consequently, with disturbed signals or in cases where cycle identification is difficult, they become unreliable. By using a log power spectrum, it is possible to transition from the time domain to the spectral domain and observe the frequency distribution of the signal's energy. However, in recent years, the focus has shifted to the cepstral domain, which involves applying a log power spectrum on the frequency domain to evaluate how the harmonic components are periodic in the spectrum.

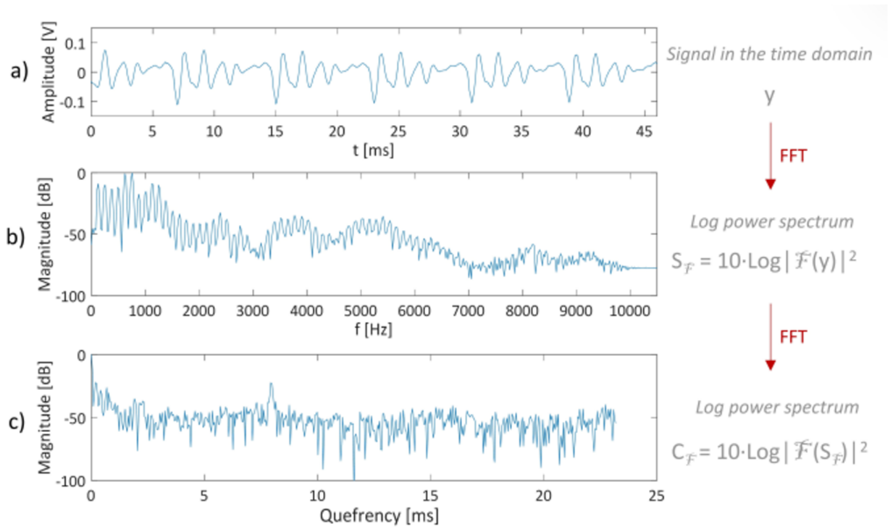


Figure 2.5: From signal in time to cepstral domain

Examples of parameters from this domain include cepstral peak prominence (CPP) and its smoothed version (CPPS). [14] These parameter, SPL and HNR are illustrated below.

Average Pitch Period

T0 is a measure of the average duration of a vocal fold vibration cycle, calculated by excluding areas where the voice exhibits interruptions or anomalies. It is expressed in milliseconds and is based on the average of the fundamental frequency periods extracted from the voice recording. It is computed as:

$$T_0 = \frac{1}{N}\sum_{i=1}^{N} T_{0_i} \tag{2.1}$$

Where $T_{0\mathrm{i}}$, i=1,2...N extracted pitch period data and N=PER-number of extracted pitch periods

Fundamental frequency

F0 is the average of all frequencies extracted from the identified periods, excluding the silent frames

$$F_0 = \frac{1}{N}\sum_{i=1}^{N} F_{0_i} \tag{2.2}$$

Where $F0_{0i} = \frac{1}{T0_{0i}}$ period to period fundamental frequency, $T_{0\mathrm{i}}$, i=1,2...N extracted pitch period data and N=PER-number of extracted pitch periods.

F0 is defined as the reciprocal of T0, so it is the inverse of the corresponding period, and the unit of measurement is Hertz (Hz). The correct extraction of this parameter is significant because many subsequent features are derived from it. Additionally, excluding the silence frames, the pre-processing activities must be accurate. If the voiced or unvoiced frames are incorrect, the fundamental frequency (F0) will also be affected. Therefore, obtaining an accurate value for F0 depends on the correct division into periods or frames, the choice of thresholds, and adherence to the conditions for jump frequency to avoid noise or artifacts. As mentioned earlier, the F0 range differs between male and female patients: males typically have values between 75 and 300 Hz, while females have values between 100 and 400 Hz. But why do females have higher values? This difference arises from anatomical and physiological characteristics between the sexes. During puberty, increased testosterone levels act on androgen receptors in the vocal folds, causing them to grow longer and thicker in boys compared to girls. Consequently, men's larger vocal folds vibrate at about half the F0 of women's during phonation. Additionally, boys' larynges descend during puberty, resulting in a longer vocal tract and lower, more closely spaced formant frequencies [15].

Jitter

- **Absolute Jitter**
  An assessment of the variability in pitch period from one cycle to the next within the analyzed voice sample, with voice break areas excluded, refers to the cycle-to-cycle temporal variation in the fundamental frequency of the voice Figure 2.6.
  $$Jita = \frac{1}{N}\sum_{i=1}^{N}|T_{0_i} - T_{0_{i+1}}| \tag{2.3}$$
  Where $T_{0\mathrm{i}}$, i=1,2...N extracted pitch period data and N=PER-number of extracted pitch periods. It is an absolute measure, expressed in microseconds, and is dependent on the average fundamental frequency of the voice. Jitter assesses the regularity and stability of a vocal signal by analyzing the variations occurring within it, the more stable signal, the lower the jitter. According to the modern literature, this parameter is analyzed when produced by computations on a sustained vowel, since for this test the patient is asked to maintain the vowel as stable as possible for as long as they can. That is the only one scenario in which the jitter can provide valuable data to identify the phonatory condition of the patient, while during the free speech it loose.

- **Jitter Percent**
  It has the same meaning and role as Jita, but it measures the relative variation of F0 rather than the absolute variation. Therefore, it is typically expressed in percentage by means of the following equation:
  $$Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_{0_i} - T_{0_{i+1}}|}{\frac{1}{N}\sum_{i=1}^{N}T_{0_i}} \tag{2.4}$$
  Where $T_{0\mathrm{i}}$, i=1,2...N extracted pitch period data and N=PER-number of extracted pitch periods.

- **High Jitter value:** Tremulous or irregular voice associated with pathological patients.

- **Low Jitter value:** Stable and regular voice associated with healthy patients.

Shimmer

- **Absolute shimmer**
  The meaning of shimmer is similar to jitter, as both analyze the regularity of

the voice signal. However, in this case, shimmer refers to the cycle-to-cycle amplitude variation in the voice's fundamental frequency.

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}|20\log(\frac{A_{i+1}}{A_i})| \tag{2.5}$$

Where A_i, i=1,2...N extracted peak to peak amplitude data and N = number of extracted impulses. It is highly sensitive to amplitude variations between consecutive pitch periods. However, errors in pitch extraction can significantly impact its value. Shimmer, which measures amplitude variation Figure 2.6, is expressed in dB. As with jitter, low values of shimmer indicate stable signals and it is taken into account while studying the sustained vowel pronunced by the patient.

- **Shimmer Percent**
  Following what is previously explained for jitter, shimmer percent has the same meaning and role as shimmer in dB, but it involves a percentage calculation rather than an absolute one as explain in the formula:

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \tag{2.6}$$

Where A_i, i=1,2...N extracted peak to peak amplitude data and N number of extracted impulses.

- **High Shimmer value:** Tremulous or irregular voice associated with pathological patients.

- **Low Shimmer value:** Stable and regular voice associated with healthy patients.

To highlight what shimmer and jitter represent and to illustrate their differences, the following figure is provided.

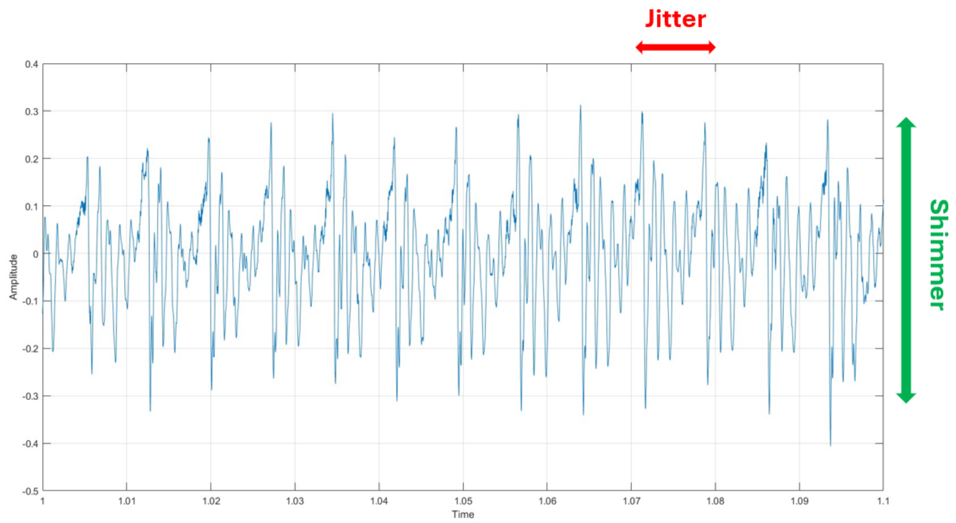


Figure 2.6: Differences Between Jitter and Shimmer

Cepstral Peak Prominence Smoothed
CPPS, Cepstral Peak Prominence Smoothed, is a parameter in the cepstral domain that includes two smoothing steps before the calculation of the CPP, which is a measure of the cepstral peak amplitude. To calculate CPPS, as previously mentioned, the process starts in the time domain, and by using two FFTs (Fast Fourier Transforms), it is possible to obtain the cepstral domain.

$$C_p = 20 \cdot \log_{10} |\mathcal{F}\{20 \cdot \log_{10}(|\mathcal{F}\{s(t)\}|)\}| \tag{2.7}$$

where C_p is the cepstrum vectror, $\mathcal{F}$ is the Fourier transform of the variable and $s(t)$ is the signal time series. In this domain, quefrency is the inverse of frequency and is expressed in milliseconds. CPPS is crucial for evaluating the regularity of the harmonic peak in the spectrum. For the Matlab R2022a implementation, there are some precautions [16]. First, the sampling rate must be set to 22050 Hz. Therefore, for sustained vowel /a/, it is necessary to resample and choose a new window length of 1024 (1024/22050 = 46 ms). To obtain CPPS, there are initially 2 smoothing steps:

- **Smoothing in time:** to reduce fluctuations and achieve a more stable signal. In this case, there are temporal frames, and instead of considering a single frame, the average of the current frame, the previous three frames, and the following three frames is evaluated, creating a mean over 7 frames. Thus, a temporal frame of 14 ms (7 frames) is considered.

- **Smoothing in cesptrum:** the concept is similar to smoothing in time, but here, instead of frames, we have bins. The cepstrum can be viewed as a series of bins along the quefrency axis. In this case, instead of using a single bin, each bin is replaced by the average of the current bin, the three previous bins, and the three following bins to achieve a smoother cepstrum along the quefrency.

Subsequently, a linear regression is calculated between 1 ms and the maximum quefrency. The 1 ms threshold is excluded because the cepstrum at low quefrencies is more affected by the spectral envelope, which varies slowly, compared to the spectral periodicity. At this point, the CPPS is evaluated as the difference in dB between the peak in the cepstrum and the corresponding value at the same quefrency on the linear regression line. A range is chosen to find the cepstral peak between 3.3 ms (300 Hz) and 16.7 ms (60 Hz) to cover all relevant frequencies for both women and men. The following image illustrates how to find CPPS in the cepstrum domain.

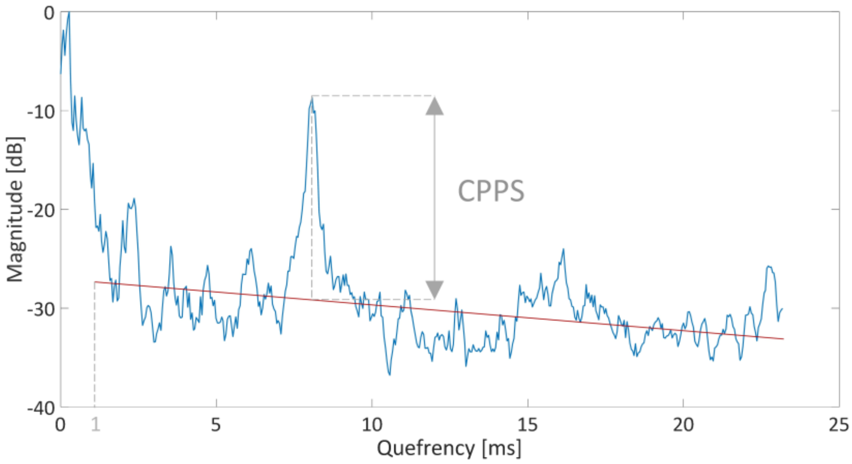


Figure 2.7: CPPS evaluation [16]

CPPS is a parameter used for both sustained vowel /a/ and free speech. It is typically analyzed with different statistics such as mean, median, standard deviation, 95th percentile, 5th percentile, kurtosis, skewness, mode, and range. All of these are essential for evaluating the behavior of the vocal signal.

- **High CPPS value:** Healthy voice.

- **Low CPPS value:** Pathological voice.

Harmonics to Noise Ratio

Harmonics to Noise Ratio is defined from the autocorrelation AC [17]. If the time signal $x(t)$ is stationary, so its statistics are constant, the autocorrelation is computed as a function of the lag τ:

$$\mathrm{AC} = \int x(t) \cdot x(t+\tau)\, dt \tag{2.8}$$

If the signal is periodic it means that outside the global maximum for $\tau = 0$ there are also global maxima outside 0, so the lag identified is call T_0, which is the period. A signal $x(t)$ can be created by combining a periodic signal $H(t)$ with noise $N(t)$. If the periodic signal and the noise are uncorrelated, the autocorrelation of $x(t)$ is the sum of the autocorrelarions of $H(t)$ and $N(t)$. When the noise is white, a peak in the autocorrelation at lag $\tau_{\max} = T_0$ indicates the power of the periodic component of the signal relative to noise. To evaluate the harmoniousness of the voice signal, the parameter HNR is calculated as:

$$\mathrm{HNR\ (in\ dB)} = 10 \log_{10} \left(\frac{\mathrm{AC}(T)}{\mathrm{AC}(0) - \mathrm{AC}(T)} \right) \tag{2.9}$$

Where $AC(T)$ is the autocorrelation at lag T and $AC(0)$ at zero lag. For perfectly periodic sounds, the HNR is infinite. For non-stationary (i.e., dynamically changing) signals, the short-term autocorrelation at a time t is estimated from a windowed segment of the signal centered around t. To obtain the relative power of the harmonic components in the numerator and the relative power of the noise in the denominator, it is necessary to normalize the autocorrelation with respect to AC(0):

$$\mathrm{HNR\ (in\ dB)} = 10 \log_{10} \left(\frac{\frac{\mathrm{AC}(T)}{\mathrm{AC}(0)}}{1 - \frac{\mathrm{AC}(T)}{\mathrm{AC}(0)}} \right) \tag{2.10}$$

HNR is evaluated on non-silent frames, and choosing a threshold based on the HNR value is essential for distinguishing between harmonic and non-harmonic frames. It is defined as the ratio of the periodic component to the non-periodic component within a segment of voiced speech. The periodic component arises from the vibration of the vocal cords, while the non-periodic component is caused by glottal noise, expressed in dB.[18] The comparison between these components reflects the efficiency of speech:

- **High HNR value:** Associated with a sonorant and harmonic voice, indicating a healthy voice.
- **Low HNR value:** Associated with an asthenic voice and dysphonia, indicating a pathological voice.

Sound Pressure Level

SPL, Sound Pressure Level, is a measure of the sound pressure of a signal relative to a reference level. It is used to quantify the intensity of sound and is expressed in decibels (dB).

$$\mathrm{SPL} = 20\log_{10}\left(\frac{p}{p_0}\right) \tag{2.11}$$

Where p is the measured sound pressure and p_0 is the reference sound pressure, typically $20\,\mu$Pa, which is the threshold of hearing for the human ear. SPL assumes different values depending on how far the recording instrument is from the sound source. The formula is based on a distance $d0$ if the recording is done at a different distance $d1$, it is necessary to adjust accordingly:

$$\mathrm{SPL}_{d1} = \mathrm{SPL}_{d0} + 20\log_{10}\left(\frac{d0}{d1}\right) \tag{2.12}$$

- **High SPL value:** Loud voice associated with healthy patients.
- **Low SPL value:** Soft voice associated with pathological voice.

Voiced/Silence

$$\frac{V}{S} = \frac{\text{voiced\_frames}}{\text{silence\_frames}} \tag{2.13}$$

This is the ratio between voiced frames and silent frames. This parameter is useful for understanding how much a person is actually speaking compared to moments of silence, so it's used only for free speech because it is not useful for vowels.

- **High Voiced/Silence value:** Good speech fluency, which is therefore associated with healthy voices.
- **Low Voices/Silence value:** Speech or voice disorder, which is therefore associated with patological voices.

Harmonic/Voiced

$$\frac{H}{V} = \frac{\text{Harmonic\_frames}}{\text{Harmonic\_frames} + \text{Unharmonic\_frames}} \tag{2.14}$$

This ratio considers only voiced frames. It represents the harmonic frames divided by the sum of harmonic and unharmonic frames. Harmonic frames represent a clean and regular voice, while unharmonic frames may indicate vocal problems or disordered phonation. Therefore, this parameter is used to assess voice quality in terms of harmonicity.

- **High Harmonic/Voiced value:** Healthy voices.
- **Low Harmonic/Voiced value:** Patological voices.

2.4 In-air microphone

The Don Carlo Gnocchi foundation gives WAV files obtaining from the microphone which are then cut properly with Audacity software, to remove the initial and final silence parts. After this initial cleaning, it is possible to import the files to be examined into Matlab. Regarding the preprocessing activities performed on these files, there are variations between the vowel and reading/monologue materials.

2.4.1 Data processing

Data processing is necessary to obtain accurate parameters, which involves cleaning the signal, choosing the correct initialization parameters, and applying all subsequent settings:

- **1° Step** Initialization of parameters, including the channel number, the threshold for distinguishing between voiced and unvoiced sounds, vertical resolution, the minimum number of frames to discard before accepting a new valid frame, and the window function along with its length.
- **2° Step** Gender of the patient is identified, as different minimum and maximum frequencies are associated with each gender:
  - **Female:** Maximum frequency 400 Hz, minimum frequency 100 Hz.
  - **Male:** Maximum frequency 300 Hz, minimum frequency 75 Hz.
- **3° Step** Voice signals obtained from an in-air microphone are stereo WAV files, so only one channel needs to be selected.
- **4° Step** The code then verifies the additive noise, ensuring it is at a low level to avoid introducing additional noise, and performs vertical resampling.
- **4° Step** The mean of the signal is removed and the signal is normalized.
- **5° Step** Silence removal is carried out by dividing the signal into frames of the chosen window length, evaluating the RMS value of each frame, and comparing it with a selected threshold to remove frames with an RMS value lower than the threshold, thus eliminating silent segments.

This part of the process is the same for both vowels and monologues/lectures, with some different values, but the code logic remains consistent. To compare VH and in-air microphones, a sampling rate of 44100 Hz and a window length of 2048 samples are used resulting in a frame length of 46 ms (2048 / 44100 Hz). Specifically for the evaluation of CPPS, the sampling rate is set to 22050 Hz and the window length is 1024 samples. Consequently, the frame length is 46 ms (1,024/22,050) also in this case.

2.4.2 Pre-processing

The pre-processing stage is crucial for distinguishing between voiced and unvoiced frames. The approach used is similar for monologues/lectures and sustained vowels, but there are significant differences between these two groups. In the case of monologues and readings, the analysis is performed on intervals corresponding to 46 ms frames that have been identified. For sustained vowels /a/, on the other hand, the identification of pseudoperiods is achieved through autocorrelation, selecting the maximum value to determine the signal's period. The pre-processing process involves different steps:

- **1° Step** Removing the mean value of the signal to center the frame around zero.

- **2° Step** Threshold based on the RMS (Root Mean Square) value is used to differentiate between silence and voice. The RMS threshold is given by:

$$RMS_{th} = \frac{rms(y)}{threshold\_div} \tag{2.15}$$

  where rms(y) is the Root Mean Square of the signal and threshold_div is initialized to 2. This means that the RMS threshold is set to half of the overall RMS of the interval. If the calculated RMS is greater than the threshold, the interval is considered to contain voice, and the analysis proceeds.

- **3° Step** To assess the harmonic quality of the frame, the HNR (Harmonic-to-Noise Ratio) is calculated. A specific HNR threshold is used to classify frames as either harmonic or non-harmonic. Here three different threshold values are provided to evaluate various results:

  – **Standard threshold: 0 dB** Harmonic and non-harmonic frames are identified correctly. However, for patients with MS, this threshold might miss some important distinctions because some frames, which are considered non-harmonic in this way, can instead be characteristic of MS.

  - **Additional thresholds: -10 dB and -20 dB** To ensure a more comprehensive analysis of the entire signal. Subsequently, it is explained how these specific thresholds are chosen.

- **4° Step** By ensuring that the frequency jump between adjacent frames or pseudoperiods is not greater than half an octave, it is possible to determine the appropriate fundamental frequency and select the non-silent harmonic signal.

2.4.3 Data qualification

At this stage, it is essential to select an appropriate HNR (Harmonics-to-Noise Ratio) threshold to distinguish frames. The process begins with data qualification, initially using a very low HNR threshold, such as -60 dB, to retain the entire signal. This allows for the evaluation of the F0 distribution to determine which frequencies should be retained or discarded. Various techniques can be employed for this evaluation, including:

- **A threshold set at the 95th percentile**
- **A threshold set at the 5th percentile**
- **A bimodal threshold**

A percentile is a statistical measure that indicates the value below which a certain percentage of data points in a dataset falls, So, in the case of the 95th percentile, it refers to the value below which 95% of the observations in the dataset fall and the same is for 5th percentile but with the value of 5%. The bimodal threshold is important because a bimodal histogram likely indicates the presence of noise or artifacts, which need to be removed. The implementation begins by applying a kernel to the fundamental frequency (F0) values to approximate the distribution, effectively estimating the shape of the histogram. Peaks are then evaluated, and if two or more are detected, the distribution is classified as bimodal. The threshold is determined by identifying the minimum value between the detected peaks.

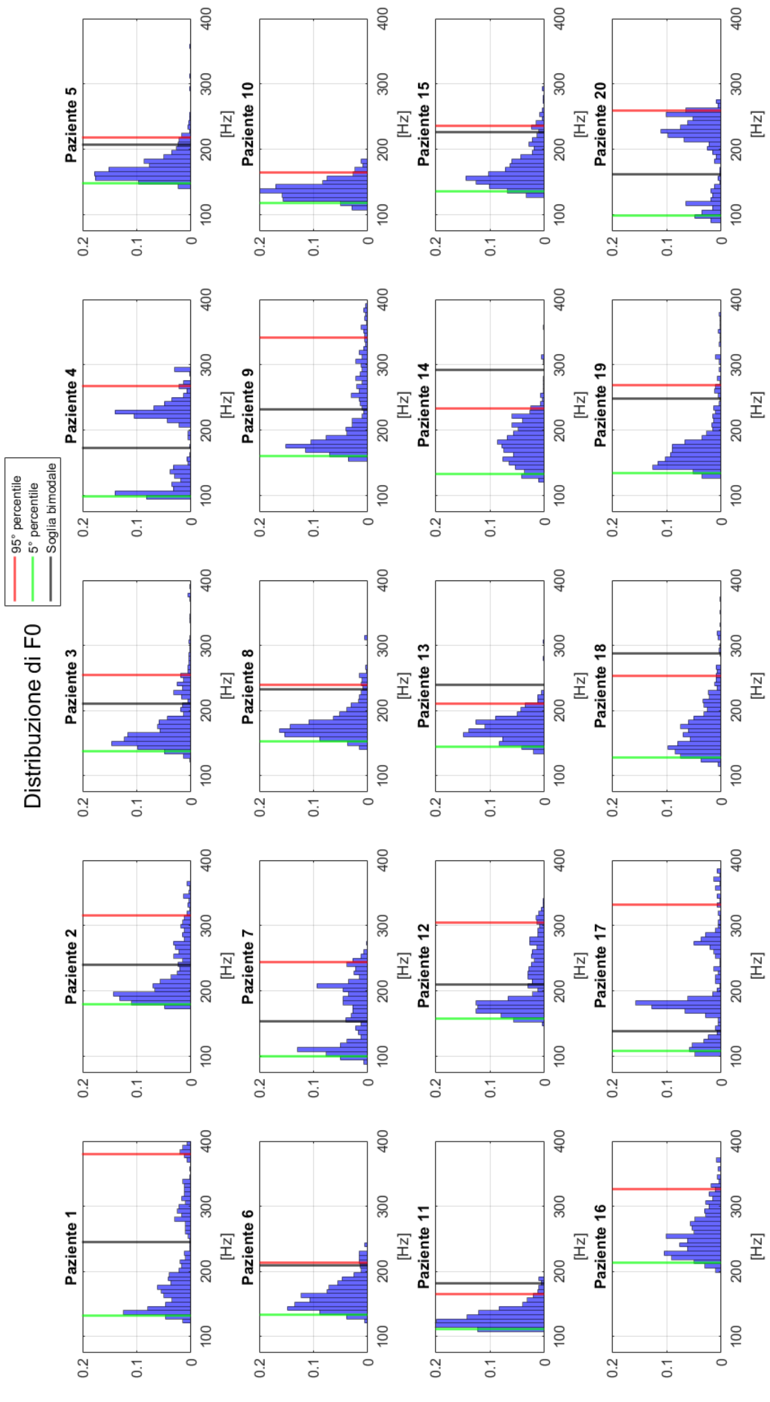


Figure 2.8: Three thresholds on F0 histogram.

From the graphs presented in Figure 2.8 the 5th percentile is not useful because it eliminates almost all frequencies, as expected from its definition. Therefore, the choice comes down to either the bimodal or the 95th percentile threshold. To determine whether the bimodal or 95th percentile threshold is more appropriate, the average difference between the initial HNR value (obtained using a -60 dB threshold) and the HNR value after applying either the 95th percentile or bimodal threshold is evaluated. In this analysis, it is found that the average difference in the bimodal case is 1.2, while in the 95th percentile case, it is 0.18. Based on this criterion, the bimodal threshold is chosen as the technique for trimming the frequency histogram. The following set of graphs is used as reference, showing the behaviour of F0 with respect to different values of HNR, such as 0 dB, -10 dB, -20 dB, -30 dB, and -40 dB.

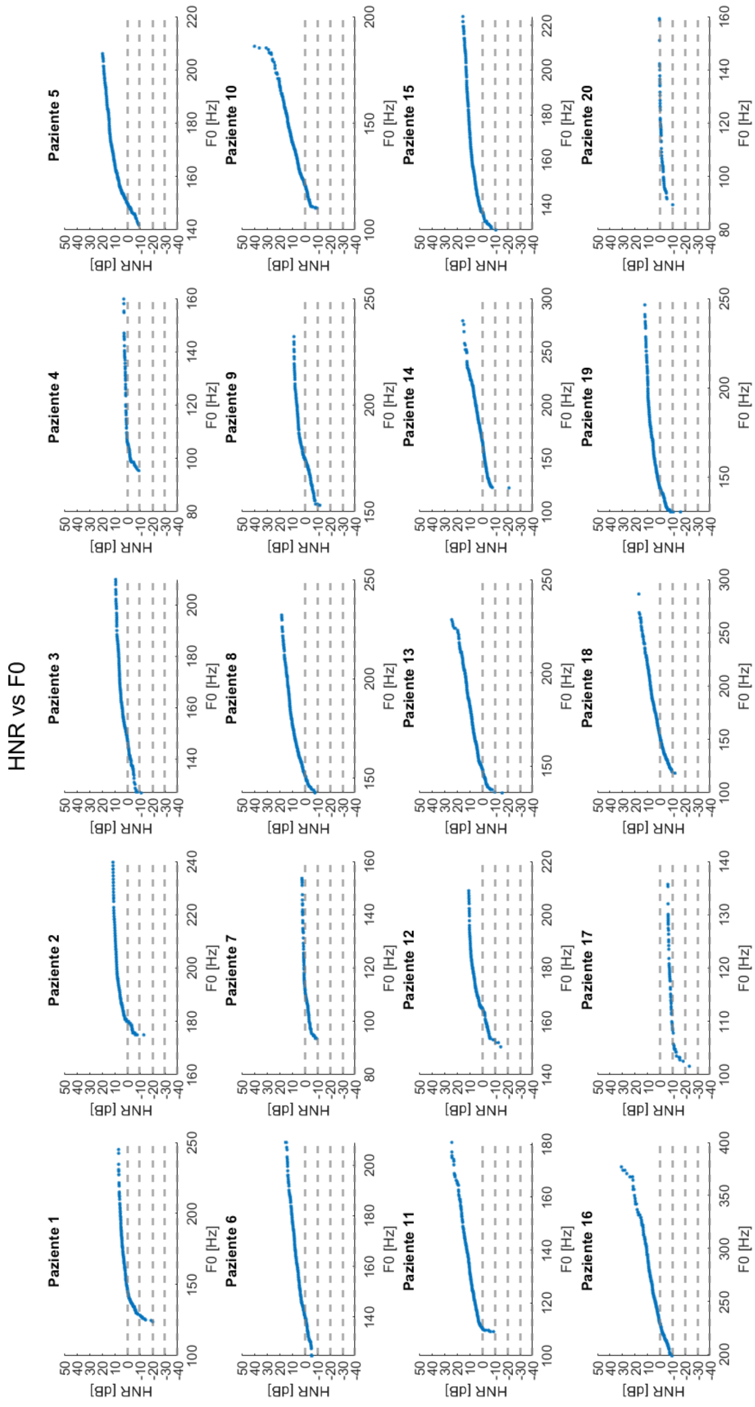


Figure 2.9: Evaluation of Different HNR Threshold Values.

The image Figure 2.9 demonstrates that the 0 dB threshold clearly distinguishes between harmonic and non harmonic frames. However, -10 dB and -20 dB thresholds can also be considered to observe how the results change. The 0 dB threshold, which correctly separates harmonic from non-harmonic frames, is not effective for healthy patients. For MS patients, including non-harmonic frames as harmonic can yield better results by highlighting additional issues with the parameters being evaluated.

2.4.4 Feature extraction

After the pre-processing phase, the extraction of the characteristic parameters takes place. The parameters extracted from the sustained vowel /a/ are different from those of the monologue and reading.

- **Sustained vowel /a/**

Parameters
Foundamental frequency (F0)
Absolut Jitter (Jita)
Jitter Percent (Jitt)
Absolute Shimmer (ShdB)
Shimmer Percent (Shim)
Cepstral Peak Prominence Smoothed (CPPS)

Table 2.2: List of sustained vowel /a/ parameters for in air microphone.

The mean value of F0 can be extracted, which is fundamental for deriving other parameters, along with the F0 vector for all the pseudoperiods found. CPPS is extracted using the statistics of mean, median, and standard deviation. As mentioned, there are three different repetitions of the vowel. It is possible to consider only one of them; typically, the second is preferred. However, in this work, the mean of the parameters from all three vowel repetitions is used.

- **Monologue and readings**

Parameters
Foundamental frequency (F0)
Harmonic to Noise Ratio (HNR)
Sound Pressure Level (SPL)
Cepstral Peak Prominence Smoothed (CPPS)

Table 2.3: List of monologue and reading parameters for in air microphone.

In this case, a vector is extracted for each parameter, representing the values within a single identified frame, in order to understand the parameters trend. Additionally, different statistics are extracted for each parameter:

Statistics
Mean
Median
Mode
Standard deviation
Range
5th percentile
95th percentile
Skewness
Kurtosis

Table 2.4: List of statistics

2.5 Vocal Holter Device

The Don Carlo Gnocchi Foundation provides TXT files obtained from the Vocal Holter (VH). To analyze these files, it is necessary to import them into Matlab R2022a to evaluate the values and compare them with the results from the in-air microphone. The files obtained from the VH and the parameters returned are different in the case of vowels compared to free speech (monologue + reading).

- **Sustained vowel /a/**

  Considering that there are three repetitions of the vowel, the VH device returns three TXT files, each containing the parameters corresponding to one repetition.

```
Short evaluation started on date: 2023-10-13 10:25:33
Calibration parameters: Mean squared error: 3.92 dB, linearity: 90 %, Intensity ratio: 0.81 : 1
Battery charge: 91 (%) Temperature: 28.3 (C) Relative humidity: 67 (%RH)


Fundamental frequency: 147.5 Hz
Jitter: 0.82 %
Shimmer: 2.78 %
CPPS (median): 17.2 dB
CPPS (standard deviation): 1.73 dB
```

Figure 2.10: TXT files from VH for sustained vowel /a/

Figure 2.10 provides an example. It initially reports the type of evaluation, whether long or short, for the vowel it's short. Additionally, it includes the calibration parameters and the environmental conditions and device status at the time of recording. Then, the evaluated parameters are reported along with their corresponding values and units of measurement. Similarly to the case with the in-air microphone, here too it is possible to use only the second repetition or to average the three repetitions, as done in this study.

Parameters
Foundamental frequency (F0)
Jitter percent (Jitt)
Shimmer percent (Shim)
Cepstral Peak Prominence Smoothed (CPPS)

Table 2.5: List of sustained vowel /a/ parameters for VH

- **Monologue and reading**

  In this case, the audio recording includes both monologue and reading within the same file, as a result, the text file from the VH contains results from both. Similarly to the vowel case, the file starts with the type of evaluation, which is long. It then reports the calibration parameters, as well as the environmental conditions and device status at the time of recording. The VH returns various files, but two of them are particularly relevant for this thesis work:

  – **"Parameters tab"** It is a table with different parameters:

Parameters
Background Noise Level (BNL)
Voicing Time Percentage (PPT)
Sound Pressure Level (SPL)
Foundamental frequency (F0)
Cepstral Peak Prominence Smoothed (CPPS)

Table 2.6: Parameters tab - monologue and reading of VH

This parameters are reported with their characteristic statistics:

Statistics
Mean
Median
5th percentile
95th percentile
Standard deviation

Table 2.7: List of statistics

- **"Parameters 46ms"**
  It reports the values, frame by frame, for Fundamental Frequency (F0) and Sound Pressure Level (SPL). This evaluation allows for understanding how these values vary across the entire signal, frame by frame. Therefore, with this Text file, it is possible to analyze the trend of these parameters and to create a vector for these two quantities.

Parameters
Foundamental frequency (F0)
Sound Pressure Level (SPL)

Table 2.8: Parameters 46 ms - monologue and reading of VH

Regarding SPL, it's important to note that the VH is positioned 22 cm from the mouth, so it is necessary to use the formula in Equation 2.12.

Chapter 3

Results

In this chapter, the results obtained from both the in-air microphone and the contact microphone are presented and analyzed. Initially, the focus is on comparing the effects of in-clinc rehabilitation with tele-rehabilitation, assessing the effectiveness of these therapeutic techniques to determine if one method is superior. Following this analysis, a comparison between the two different acquisition systems, the in-air microphone and the VH microphone, is conducted, highlighting their respective performances. The analysis covers three different time points: T0, T1, and T2, however, since the T2 data is largely incomplete, it is not currently evaluated. Therefore to evaluate the effect of the rehabilitation, only T0 and T1 are considered. Based on the explanation provided in Section 2.3 about comparing low and high parameter values, it can be noted that, given the theoretical improvement in conditions due to the rehabilitative techniques, the parameters at T1 should ideally reflect an enhancement compared to T0.

3.1 Comparison between in-clinc and tele-rehabilitation for the in-air microphone

The primary objective of the study is to evaluate the non-inferiority of tele-rehabilitation compared to traditional in-clinic rehabilitation. To reach this conclusion, the dataset is divided into two groups: 10 patients undergoing tele-rehabilitation and 10 patients receiving in-clinic rehabilitation, this division allows for a thorough comparison of all relevant parameters, enabling a deeper understanding of their respective behaviors and outcomes. By analyzing these metrics across both groups, the study aims to determine whether the effectiveness of tele-rehabilitation is comparable to that of the conventional in-clinic approach, providing valuable insights into the feasibility and potential benefits of remote healthcare solutions.

3.1.1 Sustained vowel /a/

Parameters evaluation for the sustained vowel /a/ is performed by considering the average of the results from three different repetitions, in order to ensure accuracy and consistency in the assessment. This approach is applied to all patients except for patient 14 at T0 and patient 21 at T1, as they only have two repetitions available. Regarding the use of different HNR thresholds, in this study, only a threshold of 0 dB is applied for evaluations of the vowel, alternative thresholds, such as -10 dB or -20 dB, do not provide additional insights in this context, as the analysis focuses on a single vowel sound rather than continuous speech. The effectiveness of different HNR thresholds is more relevant in longer speech samples where varying thresholds can reveal additional aspects of voice quality. In this case, where only a single vowel is analyzed, the choice of HNR threshold has a limited impact.

The extracted parameters, processed using Matlab, are summarized in the following tables, providing a clear overview of the key vocal metrics measured during the study.

Table 3.1: Values of parameters in-air microphone vowel- T0.

				T0		
ID	**F0 (Hz)**	**Jitt (%)**	**Shim (%)**	**$\mathbf{CPPS_{Mean}}$ (dB)**	**$\mathbf{CPPS_{Median}}$ (dB)**	**$\mathbf{CPPS_{std}}$ (dB)**
2	219	0,42	5,0	15,2	15,3	1,58
4	172	0,29	18,9	15,5	16,4	3,48
7	136	1,35	12,0	14,8	15,0	1,56
9	179	0,84	6,8	14,7	14,9	1,90
10	121	0,36	5,9	15,9	16,0	1,72
12	167	0,38	5,3	15,8	15,8	1,59
14	203	1,77	8,1	8,3	8,0	2,27
16	271	0,37	5,0	14,7	14,8	1,44
17	110	0,65	7,5	15,0	15,3	1,72
21	237	0,53	6,3	13,8	13,8	1,44
1	143	0,82	6,9	13,8	14,3	2,33
3	141	1,67	7,8	11,1	11,2	2,62
5	157	0,63	6,4	16,9	17,1	1,57
6	152	0,78	8,8	17,1	17,6	2,19
8	160	0,67	9,0	13,6	13,9	2,19
11	114	1,92	11,3	10,2	10,2	2,09
13	171	0,72	8,9	13,4	13,6	2,09
15	183	0,60	5,6	14,6	15,1	1,95
19	194	0,64	6,8	14,8	14,8	1,79
20	132	1,22	8,3	12,9	13,2	2,72

Table 3.2: Values of parameters in-air microphone vowel - T1. Yellow color for no effects of rehabilitation, light blue for negative effects of rehabilitation.

T1						
ID	**F0 (Hz)**	**Jitt (%)**	**Shim (%)**	**CPPS$_{\mathbf{Mean}}$ (dB)**	**CPPS$_{\mathbf{Median}}$ (dB)**	**CPPS$_{\mathbf{std}}$ (dB)**
2	215	0,21	5,0	14,4	14,4	1,56
4	146	0,23	4,6	17,2	17,3	1,46
7	112	0,28	6,7	16,3	16,5	1,76
9	179	1,59	8,3	12,5	12,9	3,03
10	158	0,20	4,6	17,5	17,6	1,38
12	191	0,32	3,7	15,8	15,9	1,55
14	294	0,79	12,3	10,8	11,1	2,17
16	252	0,34	5,0	14,5	14,6	1,82
17	105	0,49	8,9	16,1	16,3	1,72
21	220	0,35	4,2	14,2	14,3	1,39
1	150	0,50	5,7	14,4	14,5	1,89
3	141	0,83	5,3	14,9	15,2	2,10
5	161	0,38	4,2	16,8	16,9	1,35
6	129	0,44	6,9	17,2	17,3	1,36
8	205	0,23	5,0	16,4	16,5	1,48
11	117	1,26	8,4	11,5	11,8	1,94
13	189	0,71	9,2	13,1	13,3	1,76
15	216	0,27	5,3	15,6	15,7	1,41
19	190	0,51	6,6	14,0	14,0	1,91
20	188	1,50	12,0	15,6	16,0	2,40

The tables display all the metrics used to evaluate the sustained vowel at T0 and T1, highlighting the rehabilitation progress for all patients. At the top of the Tables, patients treated with tele-rehabilitation are listed, and their ID rows are shaded in gray, at the bottom, the patients who underwent in-clinic rehabilitation are shown. The expected trends for each parameter between T0 and T1 are summarized as follows:

- **Jitter (%):** T1 values lower than T0.
- **Shimmer (%):** T1 values lower than T0.
- **CPPS mean and median (dB):** T1 values higher than T0.
- **CPPS std (dB):** T1 values lower than T0.

The F0 parameter is reported for both T0 and T1, but it is not considered a key indicator of the technique's benefits, nevertheless, its value is included as it is essential for the accurate assessment of other parameters. However, the intrinsic importance of F0 should be acknowledged, as a precise evaluation of F0 is crucial for deriving the other metrics. Jitter and Shimmer absolute values are not illustrated as their values vary significantly but follow the same trend as Jitter and Shimmer percentage, to avoid redundancy, the focus here is on the percentage values. Both the mean and median of CPPS are reported to assess their alignment, the values are similar but not identical, indicating that the distribution is not Gaussian. While the distribution does not follow a perfect Gaussian curve, the trends of the mean and median values are generally similar. This comparison helps to understand whether the central tendency measures are consistent and if they provide a comparable view of the overall data distribution, despite deviations from a Gaussian pattern. In the Tables, light blue highlights indicate values that do not conform to the expected theoretical trend, while yellow highlights denote values that remain unchanged between T0 and T1. These unexpected trends can stem from various factors, such as the fact that recordings at T0 are taken in a smaller room compared to T1, where recordings occur in a larger room, or, especially for SPL, patients may move, altering the distance between their mouth and the microphone at different times, or they may fail to perform their exercises properly, or simply be having an off day. This color-coding enables a clear evaluation of the rehabilitation success rate and facilitates a comparison between the two different rehabilitation techniques, allowing for an insightful analysis of their effectiveness.

Table 3.3: Percentage of tele-rehabilitation effects in-air microphone - vowel

Measure	**Negative**	**Positive**	**None**
Jitter (%)	10%	90%	-
Shimmer (%)	30%	50%	20%
$\text{CPPS}_{\text{Mean}}$ (dB)	30%	60%	10%
$\text{CPPS}_{\text{Median}}$ (dB)	40%	60%	-
CPPS_{std} (dB)	30%	60%	10%

Table 3.4: Percentage of in-clinc-rehabilitation effects in-air microphone - vowel

Measure	**Negative**	**Positive**	**None**
Jitter (%)	10%	90%	-
Shimmer (%)	20%	80%	-
$\text{CPPS}_{\text{Mean}}$ (dB)	40%	60%	-
$\text{CPPS}_{\text{Median}}$ (dB)	40%	60%	-
CPPS_{std} (dB)	10%	90%	-

Then, doing a comparison between these results it's possible to obtain:

- **Jitter:** Both methods have similar outcomes with 90% of patients showing good effects.
- **Shimmer:** In-clinic rehabilitation shows a higher percentage of patients with good effects (80%) compared to tele-rehabilitation (50%).

- **CPPS Mean:** In-clinic rehabilitation and tele-rehabilitation have similar outcomes, with 60% of patients showing good effects and 40% showing negative effects.

- **CPPS Median:** Both methods show the same percentage of patients with good and negative effects (60% good and 40% negative).

- **CPPS Standard Deviation:** In-clinic rehabilitation appears more effective, with 90% showing good effects compared to 60% for tele-rehabilitation.

This summary provides a clear overview of the effectiveness of each rehabilitation method for different parameters, allowing for a straightforward comparison of their impacts. The results indicate that tele-rehabilitation shows good outcomes but presents some limitations in parameters such as Shimmer and CPPS standard deviation, while Jitter, CPPS mean and median are comparable to the results obtained in the clinic. In summary, in-clinic rehabilitation seems more effective in certain key parameters, but tele-rehabilitation closely matches the clinic results in several measures, which could justify its use. It is also important to note that the dataset includes only 20 patients, which allows for analysis but limits the generalizability of the results, to gain a more solid and representative understanding of the effects of tele-rehabilitation compared to in-clinic rehabilitation, studies with a larger sample size would be needed; this would help reduce the impact of individual variations and provide more reliable estimates of the actual effectiveness of both methods. To provide a visual understanding of the comparison, a graph illustrating the differences between tele-rehabilitation and in-clinic rehabilitation is presented, focusing on the parameters Jitter, Shimmer and CPPS mean along with their respective standard deviations. CPPS median is excluded from the analysis as it follows a similar trend to CPPS mean, and the focus here is on the mean values, additionally CPPS standard deviation is not included as it is not deemed a crucial parameter in this context. This graph helps in highlighting the key distinctions and similarities between the two rehabilitation methods across the selected measures.

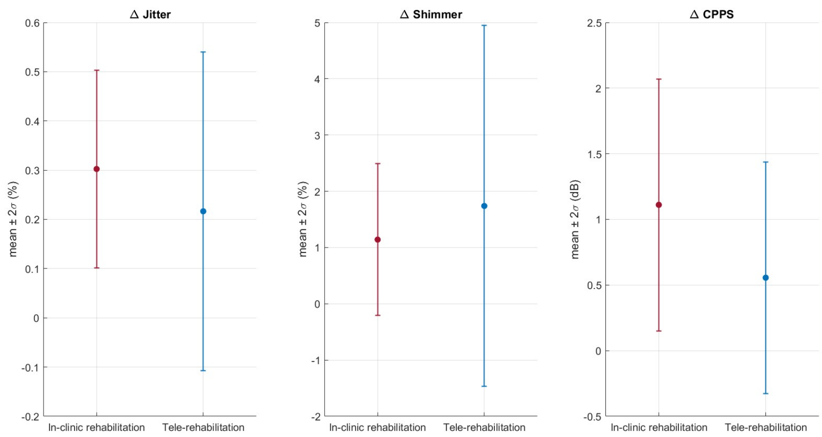


Figure 3.1: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for vowel with in-air microphone.

These Figures are derived from the initial separation of the dataset into in-clinic and tele-rehabilitation groups, the difference between T1 and T0 is then calculated for each value, taking the sign into account: for Jitter and Shimmer, since lower values are expected at T1, a negative sign is applied before calculating the difference; conversely, for CPPS, where higher values at T1 are expected, a positive sign is used. Afterward, the average and standard deviation of the resulting differences are calculated. In the Figures, the central circles represent the calculated averages, while the error bars indicate the standard deviations, which are multiplied by 2 and divided by the square root of the number of patients in each group, which is 10. So, the ordinate axis represents the confidence limits, which illustrate the confidence interval around the mean:

$$\text{Confidence Limit} = \text{Mean} \pm 2 \times \frac{\sigma}{\sqrt{n}} \tag{3.1}$$

where σ represents the standard deviation, n is the sample size and the standard deviation divided by the root mean square of the sample size is the Standard Error of the Mean (SEM). This representation shows the variability and uncertainty associated with the mean and provides a more comprehensive measure of the collected data. By examining these representations, the two rehabilitation techniques can be evaluated by checking for overlap in their confidence bands. If overlap is observed, it suggests that the two techniques are comparable; otherwise, a clear separation indicates that one technique is superior to the other. In the case presented in the Figure 3.1, this overlap is visible for each parameter, therefore it can be concluded that, for the vowel study and the specified parameters, tele-rehabilitation is comparable to in-clinic rehabilitation. This can also be highlighted by observing the data represented in the Table below, which further support the graphical findings.

Table 3.5: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for vowel with in-air microphone.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
Jitter (%)	0,22	0,51	0,30	0,32
Shimmer (%)	1,7	5,1	1,1	2,1
$\text{CPPS}_{\text{Mean}}$ (dB)	0,6	1,4	1,1	1,5

The data show that, although the average values of the analyzed parameters are similar between tele-rehabilitation and in-clinic rehabilitation, tele-rehabilitation exhibits greater variability. This is particularly evident in the higher standard deviations for Jitter, Shimmer, and CPPS mean. This suggests that, while the two methods are comparable overall, tele-rehabilitation might benefit from further optimizations to improve precision and reduce variability. Although in-clinic rehabilitation also shows high standard deviations, these values are generally lower compared to those observed in tele-rehabilitation. Therefore, while the overall performance of both methods is similar, the higher variability in tele-rehabilitation indicates that it may require more personalized protocols and more accurate monitoring to enhance results.

In conclusion, it is possible to assess that, for the sustained vowel analysis, tele-rehabilitation is not inferior to in-clinic rehabilitation, as both methods demonstrate comparable effects.

3.1.2 Reading

The reading task includes the recitation of the poem 'Notturno,' but in this analysis, only two sentences from each patient's recitation are evaluated to extract characteristic parameters. Reading parameters are evaluated for three different HNR thresholds to highlight their similarities and differences, and to assess the variations between tele-rehabilitation and in-clinic methods. Initially, the results are derived from the case where the HNR thresholds are set to 0 dB, referred to as the standard case. For simplicity and ease of comparison with other cases, this standard case is labeled as 'A', the case of -10 dB 'B' and -20 dB 'C'. The extracted parameters, processed using Matlab, are summarized in the following Tables reported only the mean values, providing a clear overview of the key vocal metrics measured during the study.

Table 3.6: Values of parameters in-air microphone reading 0 dB - T0.

T0				
ID	**HNR (dB)**	**F0 (Hz)**	**SPL (dB)**	**CPPS (dB)**
2	10,4	189	71,6	4,5
4	7,4	184	77,4	7,4
7	8,6	124	72,5	8,0
9	9,2	191	69,9	5,5
10	10,9	133	74,6	8,7
12	9,4	198	71,9	5,9
14	5,2	183	66,2	4,4
16	9,0	240	70,5	5,9
17	6,4	122	78,0	6,7
21	12,6	213	72,4	5,7
1	6,3	228	73,4	7,0
3	9,4	168	71,9	4,8
5	12,7	181	76,6	6,5
6	8,5	160	74,8	6,4
8	9,1	184	68,0	5,3
11	9,2	134	70,7	5,8
13	11,3	209	71,9	5,0
15	9,9	178	68,6	5,3
19	8,7	179	69,2	6,4
20	6,7	167	70,8	6,5

Table 3.7: Values of parameters in-air microphone reading 0 dB - T1. Yellow color for no effects of rehabilitation, light blue for negative effects of rehabilitation.

T1				
ID	**HNR (dB)**	**F0 (Hz)**	**SPL (dB)**	**CPPS (dB)**
2	13,4	224	74,0	5,2
4	11,0	127	74,5	7,4
7	10,9	129	75,1	8,6
9	9,6	191	71,3	5,5
10	9,6	145	74,6	8,1
12	12,0	226	77,6	5,5
14	8,9	180	68,4	5,3
16	10,5	261	74,5	6,1
17	8,1	128	78,4	7,7
21	12,0	218	76,4	5,5
1	10,8	171	76,8	6,2
3	9,7	139	69,5	6,8
5	11,9	174	72,5	7,1
6	9,7	169	76,7	6,6
8	11,2	201	72,8	5,5
11	11,0	137	71,4	5,6
13	11,3	212	73,6	4,8
15	11,2	188	72,6	5,6
19	9,0	187	71,0	5,8
20	6,9	182	73,7	7,4

The Tables display all the metrics used to evaluate the reading at T0 and T1, highlighting the rehabilitation progress for all patients. At the top of the tables, patients treated with tele-rehabilitation are listed, and their ID rows are shaded in gray, at the bottom, the patients who underwent in-clinic rehabilitation are shown.

The expected trends for each parameter between T0 and T1 are summarized as follows:

- **HNR (dB):** T1 values higher than T0.
- **SPL (dB):** T1 values higher than T0.
- **CPPS (dB):** T1 values higher than T0.

The F0 parameter is included in the reports for both T0 and T1, although it is not considered a primary indicator of the technique's effectiveness, its significance should not be overlooked, as precise measurement of F0 is crucial for deriving other related metrics. In the Tables, light blue shading highlights values that deviate from the anticipated theoretical trend, conversely yellow shading marks values that remain consistent between T0 and T1. These unexpected trends can stem from various factors, such as the fact that recordings at T0 are taken in a smaller room compared to T1, where recordings occur in a larger room, or, especially for SPL, patients may move, altering the distance between their mouth and the microphone at different times, or they may fail to perform their exercises properly, or simply be having an off day. This color-coding system provides a clear means to assess rehabilitation outcomes and facilitates a comparison of the effectiveness between the two rehabilitation techniques, thereby enabling a thorough evaluation of their relative success.

Table 3.8: Percentage of tele-rehabilitation effects in-air microphone - reading 0 dB

Measure	**Negative**	**Positive**	**None**
HNR (dB)	20%	80%	-
SPL (dB)	10%	80%	10%
CPPS (dB)	30%	50%	20%

Table 3.9: Percentage of in-clinc-rehabilitation effects in-air microphone - reading 0 dB

Measure	**Negative**	**Positive**	**None**
HNR (dB)	10%	80%	10%
SPL (dB)	20%	80%	-
CPPS (dB)	40%	60%	-

- **HNR:** The in-clinic rehabilitation shows a lower percentage of negative effects and a higher percentage of patients with no changes compared to tele-rehabilitation, although both methods are similar in the percentage of positive effects.

- **SPL:** Tele-rehabilitation tends to have fewer negative effects and presents some patients with no changes, whereas in-clinic rehabilitation has a higher percentage of negative effects and no cases of unchanged results.

- **CPPS:** In-clinic rehabilitation demonstrates better percentages in terms of positive effects but also has a higher rate of negative effects. Conversely, tele-rehabilitation shows a lower percentage of negative effects and a higher rate of patients with unchanged results.

These results suggest that while both methods are effective, tele-rehabilitation might be less consistent in some parameters compared to in-clinic rehabilitation, which shows a higher percentage of positive effects in some cases but also more negative effects. It is also important to note that the dataset includes only 20 patients, which allows for analysis but limits the generalizability of the results, to gain a more solid and representative understanding of the effects of tele-rehabilitation compared to in-clinic rehabilitation, studies with a larger sample size would be needed; this would help reduce the impact of individual variations and provide more reliable estimates of the actual effectiveness of both methods. To provide a visual understanding of the comparison, a graph illustrating the differences between tele-rehabilitation and in-clinic rehabilitation is presented, focusing on the parameters HNR mean, SPL mean, and CPPS mean along with their respective standard deviations. This graph helps in highlighting the key distinctions and similarities between the two rehabilitation methods across the selected measures.

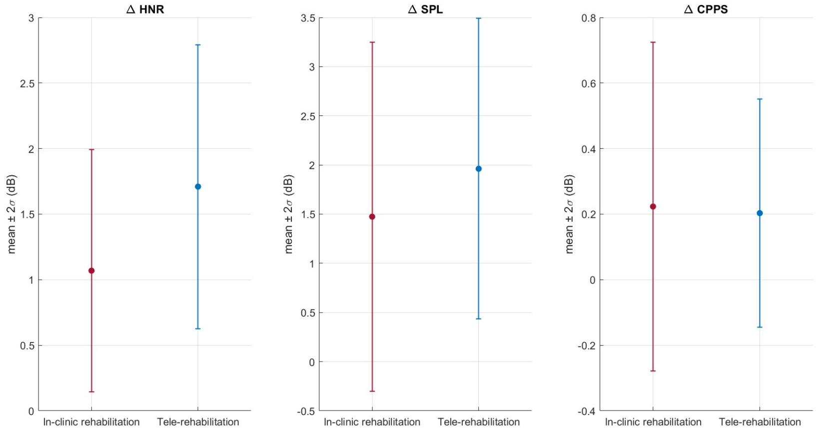


Figure 3.2: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for reading 0 dB with in-air microphone.

In the Figures, the averages of positive difference between T1 and T0 for HNR mean, SPL mean, and CPPS mean are presented, along with their average standard deviations, which are multiplied by 2 and then divided by the square root of the number of patients in each group, which is 10. The ordinate axis represents the confidence limits, which illustrate the confidence interval around the mean, as explain in 3.1.1. This representation shows the variability and uncertainty associated with the mean and provides a more comprehensive measure of the collected data. In the case presented in the Figure, the bars overlap is visible for each parameter, therefore, it can be concluded that, for the reading study in case A and the specified parameters, tele-rehabilitation is comparable to in-clinic rehabilitation. For a clearer understanding, the values shown in the graphs can be found in the following Table.

Table 3.10: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for reading 0 dB with in-air microphone.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
HNR (dB)	1,7	1,7	1,1	1,5
SPL (dB)	2,0	2,4	1,5	2,8
CPPS (dB)	0,2	0,6	0,2	0,8

From the Table, it can be observed that the average difference values for the HNR, SPL, and CPPS parameters are similar between tele-rehabilitation and in-clinic rehabilitation, indicating a comparable effect between the two methods, although HNR and SPL show slightly higher values while CPPS remains consistent across both cases under study. The average difference standard deviations, which represent the variability within the groups, show slight differences, specifically tele-rehabilitation shows greater variability for HNR and CPPS, while in-clinic rehabilitation has a higher standard deviation for SPL. Despite these differences, the overall average values and standard deviations suggest that, for the reading study analyzed, tele-rehabilitation is comparable to in-clinic rehabilitation.

Comparison of HNR thresholds

Following the analysis of the reading data in case A (with an HNR threshold of 0 dB), differences between the thresholds of -10 dB (case B) and -20 dB (case C) can be observed. As previously noted, varying HNR thresholds can lead to different results by potentially including some non-harmonic frames within harmonic ones, which affects the extracted parameter values, this discrepancy arises because the identification of harmonic frames is a crucial step in the pre-processing phase. By examining the number of identified harmonic frames and the V/S (voiced/silence) and H/V (harmonic/voiced) parameters, the following observations are obtained.

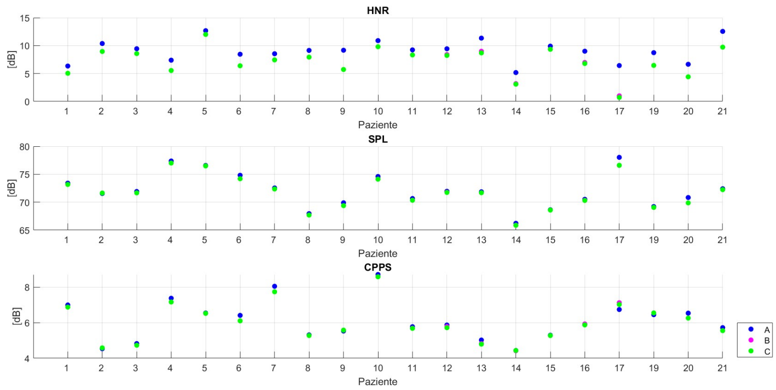
Table 3.11: Table of harmonic contents for reading T0 in case A, B and C.

ID	V/S (%)			H/V (%)			Number of har frames		
	A	B	C	A	B	C	A	B	C
2	80	80	80	67	77	77	112	130	130
4	84	84	84	30	38	38	73	96	96
7	67	67	67	51	58	58	104	117	117
9	62	62	62	47	65	65	98	138	138
10	59	59	59	80	90	90	205	229	229
12	73	73	73	70	79	80	143	159	161
14	61	61	61	37	53	54	149	215	217
16	75	75	75	65	81	82	119	148	150
17	74	74	74	16	33	34	31	68	70
21	78	78	78	69	87	87	128	160	160
1	76	76	76	39	46	46	66	78	78
3	60	60	60	53	58	58	131	140	140
5	37	37	37	76	80	80	108	114	114
6	50	50	50	54	70	70	129	164	164
8	78	78	78	69	78	78	141	159	159
11	55	55	55	73	81	81	186	206	205
13	70	70	70	62	76	78	106	131	133
15	73	73	73	73	77	77	131	139	139
19	75	75	75	57	72	72	92	118	118
20	36	36	36	34	46	46	189	252	252

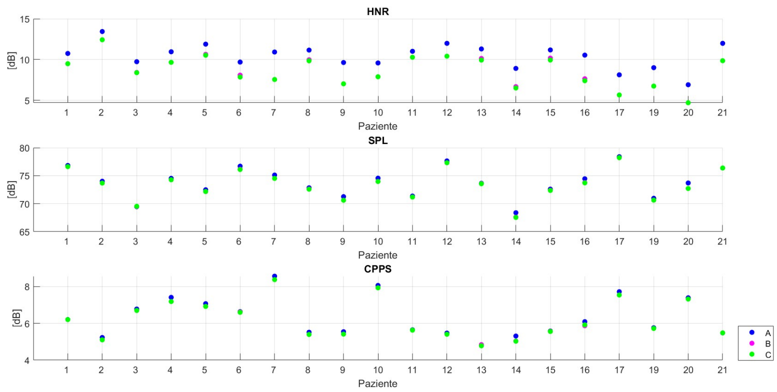
Table 3.12: Table of harmonic contents for reading T1 in case A, B and C.

ID	V/S (%)			H/V (%)			Number of har frames		
	A	B	C	A	B	C	A	B	C
2	78	78	78	85	94	94	155	171	171
4	60	60	60	63	71	71	123	139	139
7	64	64	64	36	47	47	68	88	88
9	57	57	57	57	76	76	108	143	143
10	62	62	62	73	85	85	169	198	198
12	69	69	69	82	92	92	172	194	194
14	64	64	64	57	74	75	134	173	175
16	62	62	62	67	88	90	148	196	200
17	67	67	67	41	55	55	86	118	118
21	71	71	71	66	78	78	138	160	160
1	70	70	70	59	67	67	98	111	111
3	72	72	72	56	60	60	133	143	143
5	72	72	72	74	84	84	181	205	205
6	51	51	51	66	82	83	143	174	176
8	79	79	79	73	85	86	165	191	193
11	58	58	58	78	82	82	221	233	233
13	77	77	77	82	90	91	152	168	170
15	61	61	61	73	81	82	155	173	175
19	75	75	75	61	82	82	95	127	127
20	28	28	28	40	50	50	120	149	149

In Tables the first parameter, V/S (voiced/silence), clearly appears equal for all the different thresholds because the parameter 'voice' includes the sum of harmonic and non-harmonic frames, so even if their individual numbers differ, their sum remains the same. The second parameter H/V (harmonic/voiced) shows differences between the A, B, and C cases, highlighting the distinct values between A and B and the similarities between B and C. The values in B and C are higher than in A because using a lower threshold means that frames which are considered non-harmonic in the A case are classified as harmonic in B and C. The values for B and C are equal for all patients except for patients 12, 13, 14, 16, and 17 at T0 and for patients 6, 8, 13, 14, 15 and 16 for T1, which are highlighted in light blue, where the value in case C is greater. The same trends are observed for the third parameter, which is the number of harmonic frames identified across the thresholds A, B, and C, and the differences between the parameters in cases B and C are consistent with those of H/V for all patients except one in T0, as these two parameters are correlated. Therefore, lower thresholds allow for a higher number of harmonic frames, to understand what it means to consider a higher number of harmonic frames in the cases of B and C, a graphical evaluation of the parameters HNR mean, SPL mean and CPPS mean in the different cases can be provided.



(1)



(2)

Figure 3.3: Distribution of reading parameter values comparing thresholds A, B, and C: (1) for T0 and (2) for T1.

In this figure, it is notable that case A typically shows the highest values, except for a few isolated patients, such as patients 16 and 17 at T0 for CPPS. In contrast, cases B and C generally have overlapping values, with B sometimes slightly higher, as observed in patients 12, 13, 16, and 17 for HNR at T0. A similar trend is observed at T1, where case A continues to show higher values for most patients, while cases B and C remain closely aligned, particularly for patients 5, 6, 8, 13, 14, 15, and 16 for HNR, where slight differences can be noted. This observation is further supported by the evaluation of mean values and standard deviations for 20 patients, which underline the differences in behavior between case A and cases B and C.

Table 3.13: Average values and deltas between A, B, and C for T0 reading.

Average values for 20 patients			
Parameters	**A**	Δ_{B-A}	Δ_{C-A}
HNR (dB)	9,05	-1,83	-1,88
SPL (dB)	72,05	-0,34	-0,35
CPPS (dB)	6,10	-0,07	-0,09

Table 3.14: Average values and deltas between A, B, and C for T1 reading.

Average values for 20 patients			
Parameters	**A**	Δ_{B-A}	Δ_{C-A}
HNR (dB)	10,44	-1,76	-1,83
SPL (dB)	73,77	-0,37	-0,38
CPPS (dB)	6,31	-0,10	-0,10

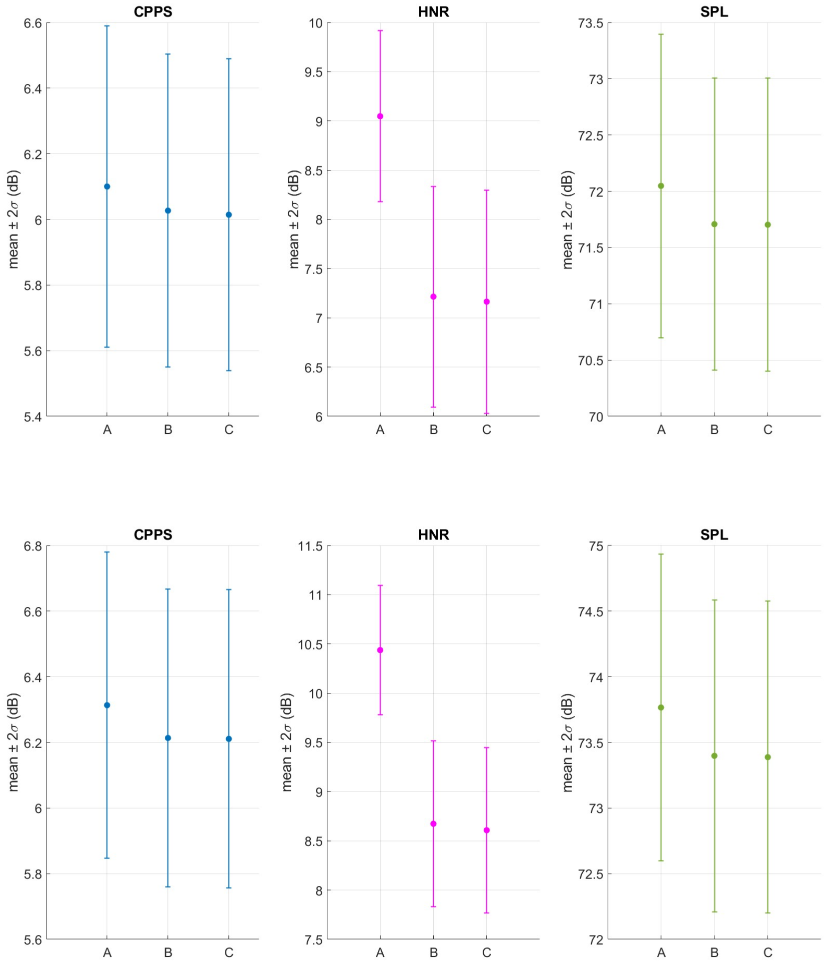
Table 3.15: Standard deviation values and deltas between A, B, and C for T0 reading.

Standard deviation values for 20 patients			
Parameters	**A**	Δ_{B-A}	Δ_{C-A}
HNR (dB)	1,99	0,58	0,61
SPL (dB)	3,10	-0,11	-0,11
CPPS (dB)	1,12	-0,03	-0,03

Table 3.16: Standard deviation values and deltas between A, B, and C for T1 reading.

Standard deviation values for 20 patients			
Parameters	**A**	Δ_{B-A}	Δ_{C-A}
HNR (dB)	1,51	0,43	0,42
SPL (dB)	2,68	0,05	0,05
CPPS (dB)	1,07	-0,03	-0,03

The Tables show that the average values decrease in both cases B and C for all parameters, while in the case of standard deviation the value increases for the HNR parameter in cases B and C, due to increased sensitivity to noise with lower HNR thresholds leading to greater variability, the standard deviation remains almost unchanged for SPL and CPPS in cases B and C.



(1)

(2)

Figure 3.4: Comparison of case A, B and C for CPPS, HNR and SPL - reading. (1) T0 and (2) T1.

The figures demonstrate that for each parameter, the use of different thresholds has comparable effects at T0, however at T1, it is evident that there is no overlap for HNR, indicating that the use of different thresholds allows for a clearer distinction in this parameter; this suggests that varying the thresholds can lead to different interpretations of the data, particularly for HNR, which may provide more reliable insights into vocal quality compared to CPPS and SPL.

To recap the results: HNR, SPL, and CPPS are the key parameters for assessment, evaluated at three different thresholds: 0 dB, -10 dB, and -20 dB. Lower thresholds result in lower values due to the inclusion of non-harmonic frames within the harmonic frame, this is a notable consideration, as non-harmonic voice frames may be significant in individuals affected by MS. This first analysis aims to evaluate the differences between the various thresholds by considering the entire dataset, which includes all 20 patients without distinguishing between in-clinic and tele-rehabilitation.

To investigate the impact of using different thresholds in relation to the comparison between the two rehabilitation methods, the following graphs are presented.

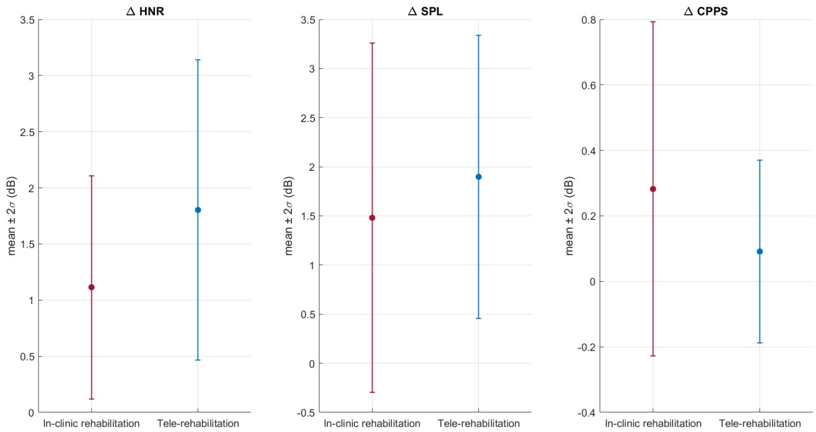


Figure 3.5: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for reading at -10 dB with in-air microphone.

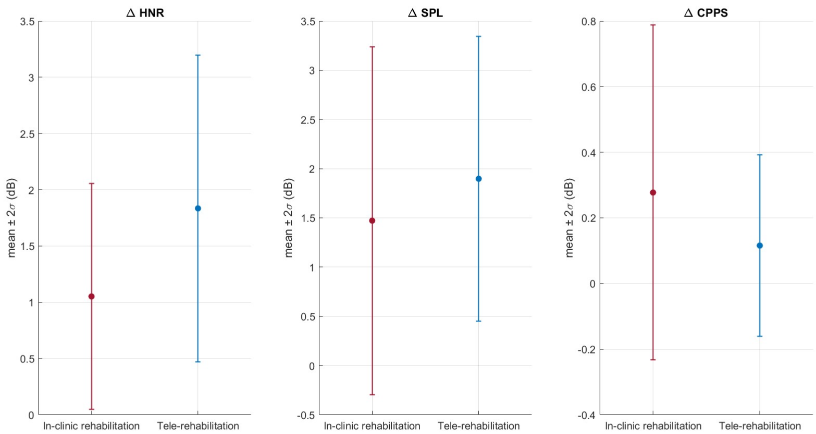


Figure 3.6: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for reading at -20 dB with in-air microphone.

The Figures illustrate that even when using different thresholds of -10 dB and

-20 dB, in-clinic and tele-rehabilitation show comparable effects. This conclusion is supported by the visible overlap of the bars for each parameter, indicating that the variability between the two rehabilitation methods is minimal. Additionally, the Tables provide a detailed breakdown of the average difference between T1 and T0 for both tele-rehabilitation and in-clinic rehabilitation, as well as the average standard deviation for each group, further supporting the graphical findings.

Table 3.17: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for reading -10 dB with in-air microphone.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
HNR (dB)	1,8	2,1	1,1	1,6
SPL (dB)	1,9	2,3	1,5	2,8
CPPS (dB)	0,1	0,4	0,3	0,8

Table 3.18: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for reading -20 dB with in-air microphone.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
HNR (dB)	1,8	2,2	1,1	1,6
SPL (dB)	1,9	2,3	1,5	2,8
CPPS (dB)	0,1	0,4	0,3	0,8

The Tables show that, for both the -10 dB and -20 dB thresholds, the average values of HNR, SPL, and CPPS are comparable between tele-rehabilitation and in-clinic rehabilitation, however tele-rehabilitation exhibits greater variability for HNR, while in-clinic rehabilitation shows a higher standard deviation for SPL and CPPS. Overall, the results indicate that the two methods produce similar effects, regardless of the threshold used. Using different thresholds, the comparison between in-clinic and tele-rehabilitation remains the same. However, to more effectively assess the differences between cases A, B, and C, it is possible to evaluate the difference between cases B and A, and C and A, in order to gain a more complete understanding.

Table 3.19: Comparison of average values and standard deviations between tele-rehabilitation and in-clinic rehabilitation for case A and B - reading.

	Tele-rehabilitation				**In-clinic rehabilitation**			
	$\overline{x_{T1-T0}}$		$\overline{\sigma_{T1-T0}}$		$\overline{x_{T1-T0}}$		$\overline{\sigma_{T1-T0}}$	
	A	Δ_{B-A}	**A**	Δ_{B-A}	**A**	Δ_{B-A}	**A**	Δ_{B-A}
HNR (dB)	1,7	-0,1	1,7	-0,4	1,1	0,0	1,5	-0,1
SPL (dB)	2,0	0,5	2,4	0,1	1,5	0,0	2,8	0,0
CPPS (dB)	0,2	0,1	0,6	0,1	0,2	-0,1	0,8	0,0

Table 3.20: Comparison of average values and standard deviations between tele-rehabilitation and in-clinic rehabilitation for case A and C - reading.

	Tele-rehabilitation				**In-clinic rehabilitation**			
	$\overline{x_{T1-T0}}$		$\overline{\sigma_{T1-T0}}$		$\overline{x_{T1-T0}}$		$\overline{\sigma_{T1-T0}}$	
	A	Δ_{B-A}	**A**	Δ_{B-A}	**A**	Δ_{B-A}	**A**	Δ_{B-A}
HNR (dB)	1,7	0,0	1,7	0,0	1,1	-0,1	1,5	0,0
SPL (dB)	2,0	0,4	2,4	0,0	1,5	0,0	2,8	0,0
CPPS (dB)	0,2	0,0	0,6	0,0	0,2	0,0	0,8	0,0

Tele-rehabilitation appears to be more effective in enhancing vocal parameters like SPL and shows more stability in maintaining or improving voice quality (HNR and CPPS). This might be due to the flexibility, increased patient engagement, and potential for more frequent practice sessions in a home setting. In contrast, in-clinic rehabilitation is more stable overall, but with fewer significant improvements, suggesting that while it may help maintain vocal abilities, it might be less effective in driving substantial improvements over time.

3.1.3 Monologue

The monologue task is 60 seconds of free speech, its parameters are extracted for the three different HNR thresholds, case A, B and C, to emphasize their similarities and differences, in the same way of reading. The parameters obtained from the study and analyzed using Matlab are presented in the following Tables, which exclusively display the average values of the key vocal metrics, offering a comprehensive summary and insight into the principal vocal characteristics assessed throughout the research.

Table 3.21: Values of parameters in-air microphone monologue 0 dB - T0. Yellow color for no effects of rehabilitation, light blue for negative effects of rehabilitation.

T0				
ID	**HNR (dB)**	**F0 (Hz)**	**SPL (dB)**	**CPPS (dB)**
2	9,8	219	71,8	5,0
4	6,3	169	75,7	7,0
7	6,7	147	70,6	7,2
9	8,1	200	70,8	5,7
10	9,0	137	73,2	8,2
12	10,3	203	71,4	6,0
14	5,2	177	67,6	5,1
16	7,1	254	70,2	5,5
17	7,0	127	76,7	7,3
21	5,0	187	73,2	7,0
1	6,2	180	71,2	5,7
3	6,8	173	69,2	5,4
5	11,9	171	72,1	6,4
6	7,5	162	73,1	5,8
8	9,7	179	68,2	5,4
11	9,0	130	73,0	6,1
13	8,7	174	69,9	5,3
15	9,2	169	70,1	6,6
18	6,8	165	66,9	5,3
19	6,8	172	67,0	6,6
20	6,8	172	67,0	6,6

Table 3.22: Values of parameters in-air microphone monologue 0 dB - T1. Yellow color for no effects of rehabilitation, light blue for negative effects of rehabilitation.

T1				
ID	**HNR (dB)**	**F0 (Hz)**	**SPL (dB)**	**CPPS (dB)**
2	11,4	227	73,4	5,3
4	9,9	114	72,6	7,8
7	8,7	152	74,0	8,5
9	8,4	179	70,4	5,6
10	8,2	136	73,6	8,2
12	12,2	231	75,4	6,3
14	7,9	169	69,3	5,8
16	8,6	269	74,0	5,6
17	7,0	125	77,1	7,5
21	11,3	220	75,7	5,2
1	10,6	161	74,2	6,3
3	8,7	168	70,3	6,4
5	10,5	174	72,4	6,6
6	7,6	144	72,8	8,1
8	10,8	195	71,4	5,4
11	9,1	126	71,8	5,9
13	10,2	193	73,2	4,8
15	8,8	174	71,2	6,4
19	7,3	197	70,1	6,1
20	8,3	143	73,1	7,6

The tables illustrate the metrics assessed at T0 and T1, focusing on the rehabilitation progress for all participants. The upper sections of the tables feature patients who received tele-rehabilitation, with their IDs marked in gray for distinction, conversely the lower sections display patients who participated in in-clinic rehabilitation. A summary of the anticipated trends for each parameter between T0 and T1 is provided below:

- **HNR (dB):** T1 values higher than T0.
- **SPL (dB):** T1 values higher than T0.
- **CPPS (dB):** T1 values higher than T0.

The F0 parameter is reported for both T0 and T1, and although it is not considered a primary measure of the technique's effectiveness, its importance should not be overlooked, as accurate F0 measurement is crucial for deriving other related metrics. In the Tables, values that diverge from the expected theoretical trend are highlighted with light blue shading, in contrast yellow shading indicates values that are the same between T0 and T1. These unexpected trends can stem from various factors, such as the fact that recordings at T0 are taken in a smaller room compared to T1, where recordings occur in a larger room, or, especially for SPL, patients may move, altering the distance between their mouth and the microphone at different times, or they may fail to perform their exercises properly, or simply be having an off day. This color-coding approach offers a straightforward way to evaluate rehabilitation results and compare the effectiveness of the two rehabilitation techniques, allowing for a comprehensive assessment of their relative success.

Table 3.23: Percentage of tele-rehabilitation effects in-air microphone - monologue 0 dB.

Measure	**Negative**	**Positive**	**None**
HNR (dB)	10%	80%	10%
SPL (dB)	20%	80%	-
CPPS (dB)	20%	70%	10%

Table 3.24: Percentage of in-clinc-rehabilitation effects in-air microphone - reading 0 dB.

Measure	**Negative**	**Positive**	**None**
HNR (dB)	20%	80%	-
SPL (dB)	30%	70%	-
CPPS (dB)	20%	80%	-

- **HNR (dB):** Both tele-rehabilitation and in-clinic rehabilitation show comparable outcomes, with 80% of patients exhibiting positive effects. The percentage of patients experiencing negative effects is slightly higher in in-clinic rehabilitation (20% compared to 10% in tele-rehabilitation), but overall, the positive effects are consistent across both methods.

- **SPL (dB):** Tele-rehabilitation demonstrates a higher success rate with 80% of patients showing positive effects, compared to 70% in-clinc rehabilitation. The percentage of patients experiencing negative effects is also higher in in-clinic rehabilitation (30%) compared to tele-rehabilitation (20%).

- **CPPS (dB):** In-clinic rehabilitation shows a higher percentage of positive effects, with 80% of patients experiencing improvement compared to 70% for tele-rehabilitation, both methods have the same percentage of negative effects (20%). However, 10% of patients in the tele-rehabilitation group show no effect, while there are no cases with no effect in the in-clinic rehabilitation group.

Overall, both tele-rehabilitation and in-clinic rehabilitation are effective, but tele-rehabilitation generally provides slightly better outcomes for SPL and demonstrates fewer cases of no effect, in contrast, in-clinic rehabilitation excels in CPPS with a higher rate of positive effects. These findings suggest that while tele-rehabilitation may be advantageous in certain areas, in-clinic rehabilitation offers notable benefits in others, highlighting the importance of selecting the appropriate method based on specific rehabilitation goals and patient needs. It is crucial to acknowledge that the dataset comprises only 20 patients, which, while sufficient for preliminary analysis, restricts the generalizability of the findings. To obtain a more robust and representative assessment of tele-rehabilitation versus in-clinic rehabilitation, larger studies are necessary. To provide a visual understanding of the comparison, a graph illustrating the differences between tele-rehabilitation and in-clinic rehabilitation is presented, focusing on the parameters HNR mean, SPL mean, and CPPS mean along with their respective standard deviations. This graph helps in highlighting the key distinctions and similarities between the two rehabilitation methods across the selected measures.

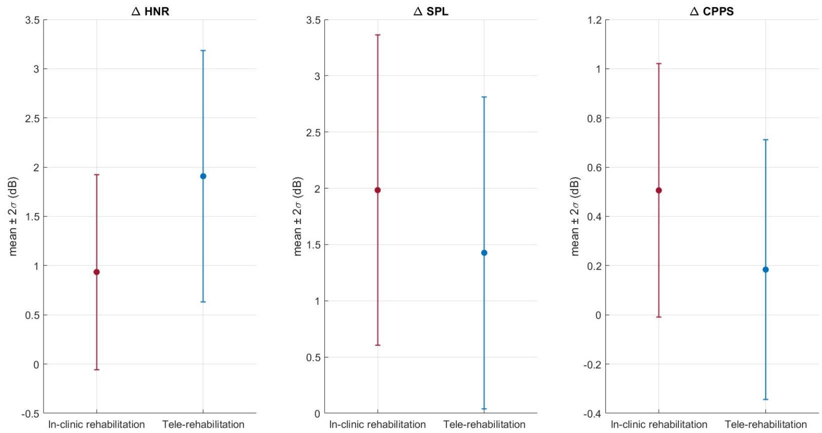


Figure 3.7: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for monologue 0 dB with in-air microphone.

In the Figures, the averages of positive difference between T1 and T0 for HNR mean, SPL mean, and CPPS mean are presented, along with their average standard deviations, which are multiplied by 2 and then divided by the square root of the number of patients in each group, which is 10. The ordinate axis represents the confidence limits, which illustrate the confidence interval around the mean, as explain in 3.1.1. This representation shows the variability and uncertainty associated with the mean and provides a more comprehensive measure of the collected data. In the case presented in the figure, the bars overlap is visible for each parameter, therefore, it can be concluded that, for the reading study in case A and the specified parameters, tele-rehabilitation is comparable to in-clinic rehabilitation. For a clearer understanding, the values shown in the graphs can be found in the following Table.

Table 3.25: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for monologue 0 dB with in-air microphone.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
HNR (dB)	1,9	2,0	0,9	1,6
SPL (dB)	1,4	2,2	2,0	2,2
CPPS (dB)	0,2	0,8	0,5	0,8

From the Table, it can be observed that the average difference values for HNR and SPL are higher in tele-rehabilitation compared to in-clinic rehabilitation, while for CPPS, the values are nearly identical for both methods. The standard deviations of the average differences are higher across both cases, indicating a significant variability in the results. Despite these differences, the data suggest that for the monologue study with HNR threshold at 0 dB, tele-rehabilitation offers comparable outcomes to in-clinic rehabilitation.

Comparison of HNR thresholds

In the analysis of the monologue data for case A, using an HNR threshold of 0 dB, distinct differences emerge when compared to thresholds of -10 dB and -20 dB, respectively case B and C. As mentioned earlier, varying HNR thresholds can yield different outcomes by potentially misclassifying non-harmonic frames as harmonic ones, which, in turn, influences the values of the extracted parameters. This variation occurs because accurately identifying harmonic frames is a critical step during the pre-processing stage. By analyzing both the number of identified harmonic frames and the V/S and H/V parameters, the following insights were obtained.

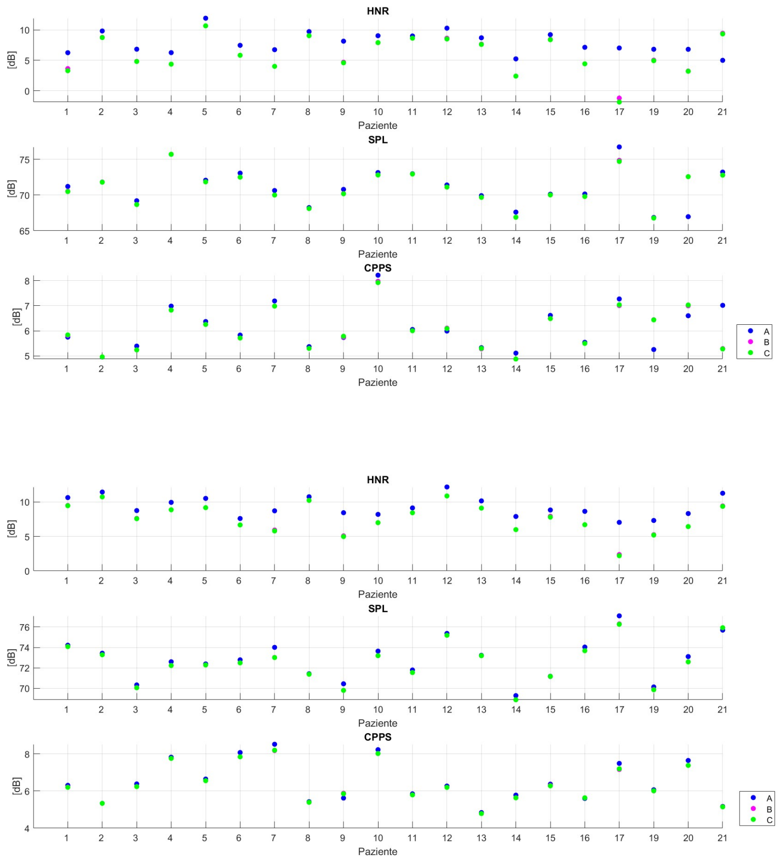
Table 3.26: Table of harmonic contents for monologue T0 in case A, B and C.

ID	**V/S (%)**			**H/V (%)**			**Number of har frames**		
	A	B	C	A	B	C	A	B	C
2	78	78	78	72	80	81	712	794	796
4	87	87	87	25	33	33	272	363	363
7	68	68	68	33	46	46	282	394	394
9	63	63	63	42	62	63	334	490	496
10	56	56	56	79	91	91	553	638	641
12	71	71	71	63	73	73	564	652	656
14	71	71	71	33	59	59	295	523	523
16	66	66	66	57	85	85	474	708	708
17	73	73	73	10	38	41	95	352	376
21	52	62	62	34	86	86	226	674	679
1	76	76	76	29	40	42	275	388	400
3	81	81	81	37	50	50	378	505	505
5	62	62	62	73	81	81	575	639	639
6	55	55	55	59	73	73	406	506	506
8	75	75	75	76	82	82	717	775	775
11	48	48	48	75	79	79	455	481	481
13	69	69	69	70	80	80	616	697	701
15	66	66	66	72	79	80	603	664	666
19	62	60	60	37	75	75	295	564	568
20	60	52	52	59	47	47	441	307	309

Table 3.27: Table of harmonic contents for monologue T1 in case A, B and C.

ID	V/S (%)			H/V (%)			Number of har frames		
	A	B	C	A	B	C	A	B	C
2	72	72	72	86	92	92	788	834	834
4	62	62	62	55	62	62	436	490	490
7	61	61	61	39	52	52	296	396	398
9	64	64	64	51	77	77	415	619	623
10	62	62	62	73	87	87	572	686	686
12	70	70	70	82	92	92	731	814	814
14	60	60	60	59	75	75	313	397	397
16	61	61	61	68	87	87	523	669	669
17	70	70	70	26	46	46	228	406	410
21	67	67	67	69	81	81	590	685	687
1	60	60	60	66	74	74	497	553	555
3	68	68	68	64	72	73	553	624	626
5	63	63	63	71	80	80	564	637	637
6	54	54	54	65	74	74	442	504	504
8	75	75	75	84	89	89	796	846	846
11	64	64	64	70	77	77	560	618	618
13	69	69	69	80	90	90	701	788	788
15	64	64	64	74	81	82	601	658	664
19	63	63	63	52	67	67	412	530	530
20	53	53	53	43	52	52	288	350	350

In Tables the first parameter, V/S (voiced/silence), clearly appears equal for all the different thresholds because the parameter 'voice' includes the sum of harmonic and non-harmonic frames, so even if their individual numbers differ, their sum remains the same. The second parameter H/V (harmonic/voiced) shows differences between the A, B, and C cases, highlighting the distinct values between A and B and the similarities between B and C. The values in B and C are higher than in A because using a lower threshold means that frames which are considered non-harmonic in the A case are classified as harmonic in B and C. The values for B and C are equal for all patients except for patients 1, 2, 9, 15 and 17 at T0 and for patient 3 for T1, which are highlighted in light blue, where the value in case C is greater. For the third parameter there are more difference the H/V, in this case the differences are for patients 1, 2, 9, 10, 12, 13, 15, 17, 19, 20 and 21 at T0, and for patients 1, 3, 4, 7, 15, 16 and 17 at T1. Therefore, lower thresholds allow for a higher number of harmonic frames, to understand what it means to consider a higher number of harmonic frames in the cases of B and C, a graphical evaluation of the parameters HNR mean, SPL mean and CPPS mean in the different cases can be provided.



(1)

(2)

Figure 3.8: Distribution of monologue parameter values comparing thresholds A, B, and C: (1) for T0 and (2) for T1.

In this Figure, it is notable that case A typically shows the highest values, except for a few isolated patients, such as patient 21 for HNR and patient 20 for SPL and CPPS at T0. In contrast, cases B and C generally have overlapping values. Similar trend is observed at T1, where case A continues to show higher values for most patients, while cases B and C remain closely aligned. This observation is further supported by the evaluation of mean values and standard deviations for 20 patients, which underline the differences in behavior between case A and cases B and C.

Table 3.28: Average values and deltas between A, B, and C for T0 monologue.

Average values for 20 patients			
Parameters	**A**	Δ_{B-A}	Δ_{C-A}
HNR (dB)	7,86	-1,84	-1,92
SPL (dB)	71,10	-0,11	-0,12
CPPS (dB)	6,13	-0,09	-0,09

Table 3.29: Average values and deltas between A, B, and C for T1 monologue.

Average values for 20 patients			
Parameters	**A**	Δ_{B-A}	Δ_{C-A}
HNR (dB)	9,29	-1,62	-1,66
SPL (dB)	72,81	-0,30	-0,30
CPPS (dB)	6,47	-0,10	-0,10

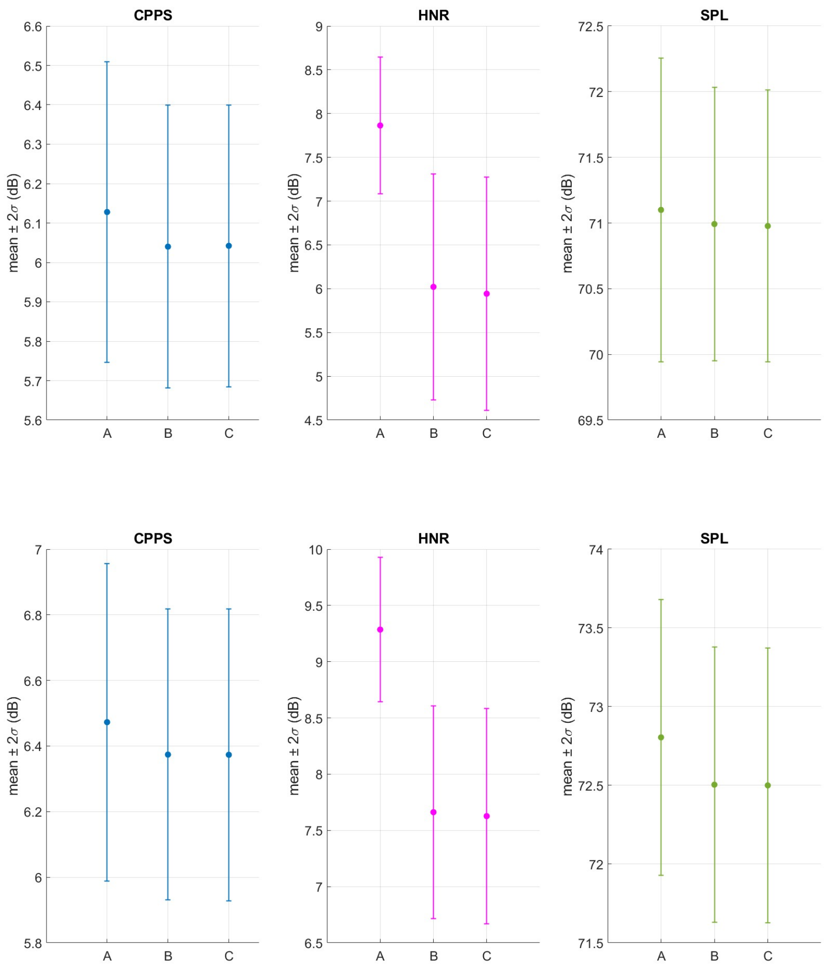
Table 3.30: Standard deviation values and deltas between A, B, and C for T0 monologue.

Standard deviation values for 20 patients			
Parameters	**A**	Δ_{B-A}	Δ_{C-A}
HNR (dB)	1,79	1,18	1,27
SPL (dB)	2,65	-0,27	-0,28
CPPS (dB)	0,87	-0.05	-0,05

Table 3.31: Standard deviation values and deltas between A, B, and C for T1 monologue.

Standard deviation values for 20 patients			
Parameters	**A**	Δ_{B-A}	Δ_{C-A}
HNR (dB)	1,47	0,69	0,73
SPL (dB)	2,01	0,00	0,00
CPPS (dB)	1,11	-0,09	-0,09

The data show a general decrease in average values across all parameters in cases B and C, however the standard deviation of the HNR parameter rises in these cases, indicating increased noise sensitivity and greater variability as the HNR threshold lowers. On the other hand, the standard deviation for SPL and CPPS remains relatively stable, with the exception of SPL at T0, where a decline in variation is observed compared to case A. In summary, HNR, SPL, and CPPS are key metrics tested at three thresholds: 0 dB, -10 dB, and -20 dB. At T1, most parameters show higher values than at T0, except for a few outliers. Lower thresholds result in smaller values as they factor in non-harmonic frames, which is particularly important in assessing individuals with MS, where non-harmonic voice segments may play a significant role. This initial analysis explores the differences across various thresholds by examining the entire dataset, which includes all 20 patients without differentiating between in-clinic and tele-rehabilitation.



(1)

(2)

Figure 3.9: Comparison of case A, B and C for CPPS, HNR and SPL - monologue. (1) T0 and (2) T1.

The Figures indicate that for each parameter, employing different thresholds yields similar effects at T0, however, at T1, it is clear that there is no overlap for HNR. This suggests that using various thresholds facilitates a clearer differentiation for this parameter. Consequently, adjusting the thresholds may result in diverse interpretations of the data, especially for HNR, which could offer more dependable insights into vocal quality than CPPS and SPL. To assess how these thresholds impact the comparison between the two rehabilitation methods, the following graphs are presented.

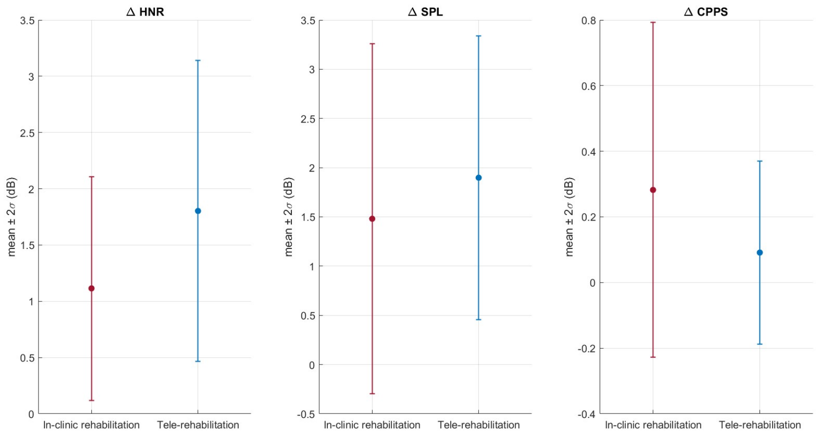


Figure 3.10: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for monologue at -10 dB with in-air microphone.

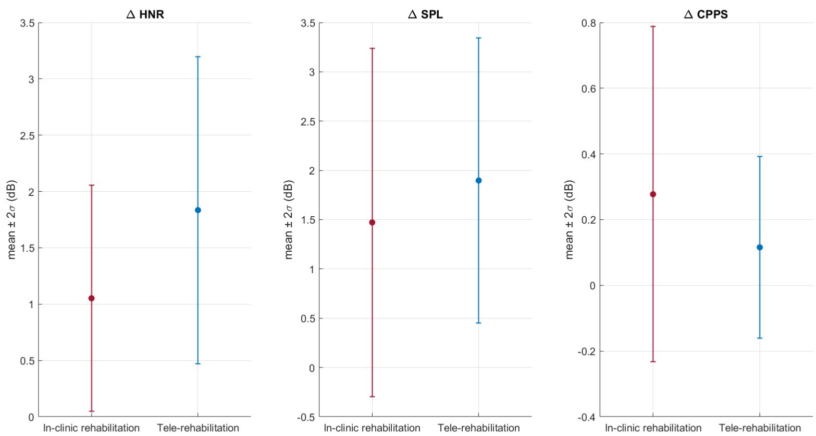


Figure 3.11: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for monologue at -20 dB with in-air microphone.

The Tables illustrate that even when using different thresholds of -10 dB and -20 dB, in-clinic and tele-rehabilitation show comparable effects. This conclusion is supported by the visible overlap of the bars for each parameter, indicating that the variability between the two rehabilitation methods is minimal. Additionally, the Tables detail the average changes between T1 and T0 for both tele-rehabilitation and in-clinic rehabilitation, along with the same average difference of standard deviation for each group. These findings complement the graphical data shown in the Figures, with each circle and bar representing the respective data points.

Table 3.32: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for monologue -10 dB with in-air microphone.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
HNR (dB)	1,9	1,7	1,3	2,1
SPL (dB)	1,5	2,3	1,5	1,8
CPPS (dB)	0,4	0,4	0,3	0,8

Table 3.33: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for monologue -20 dB with in-air microphone.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
HNR (dB)	2,0	1,8	1,4	2,2
SPL (dB)	1,6	2,3	1,5	1,8
CPPS (dB)	0,4	0,4	0,3	0,8

The Figures demonstrate that both tele-rehabilitation and in-clinic rehabilitation yield similar effects, regardless of whether the thresholds are set at -10 dB or -20 dB. This conclusion is reinforced by the overlap observed in the bar graphs for each parameter, suggesting minimal variability between the two rehabilitation methods. The accompanying Tables provide a quantitative comparison of the average changes and standard deviations from T0 to T1 for both rehabilitation methods, segmented by the two different thresholds. The results show that tele-rehabilitation and in-clinic rehabilitation have comparable average changes across all parameters (HNR, SPL, CPPS), the standard deviations however, differ, with in-clinic rehabilitation displaying slightly higher variability for HNR and CPPS. However, to more effectively assess the differences between cases A, B, and C, it is possible to evaluate the difference between cases B and A, and C and A, in order to gain a more complete understanding.

Table 3.34: Comparison of average values and standard deviations between tele-rehabilitation and in-clinic rehabilitation for case A and B - monologue.

	Tele-rehabilitation				In-clinic rehabilitation			
	$\overline{x_{T1-T0}}$		$\overline{\sigma_{T1-T0}}$		$\overline{x_{T1-T0}}$		$\overline{\sigma_{T1-T0}}$	
	A	Δ_{B-A}	**A**	Δ_{B-A}	**A**	Δ_{B-A}	**A**	Δ_{B-A}
HNR (dB)	1,9	0,0	2,0	-0,3	0,9	0,4	1,6	0,6
SPL (dB)	1,4	0,1	2,2	0,1	2,0	-0,5	2,2	-0,4
CPPS (dB)	0,2	0,2	0,8	-0,4	0,5	-0,2	0,8	0,0

Table 3.35: Comparison of average values and standard deviations between tele-rehabilitation and in-clinic rehabilitation for case A and C - monologue.

	Tele-rehabilitation				In-clinic rehabilitation			
	$\overline{x_{T1-T0}}$		$\overline{\sigma_{T1-T0}}$		$\overline{x_{T1-T0}}$		$\overline{\sigma_{T1-T0}}$	
	A	Δ_{B-A}	**A**	Δ_{B-A}	**A**	Δ_{B-A}	**A**	Δ_{B-A}
HNR (dB)	1,9	0,1	2,0	-0,3	0,9	0,4	1,6	0,6
SPL (dB)	1,4	0,1	2,2	0,1	2,0	-0,5	2,2	-0,4
CPPS (dB)	0,2	0,2	0,8	-0,4	0,5	-0,2	0,8	0,0

Overall, the results show that the differences between tele-rehabilitation and in-clinic rehabilitation are relatively minimal concerning the average changes in HNR, SPL, and CPPS parameters between cases A and B or A and C. However, the variability (standard deviation) appears to be more stable for tele-rehabilitation compared to in-clinic rehabilitation for some parameters. This suggests that, while the average values of the parameters remain similar between the two methods, the variability of the data may differ slightly, which could have implications for therapy customization and progress monitoring.

3.2 Comparison between in-clinc and tele-rehabilitation for the Vocal Holter device

The Vocal Holter generates text files that are processed in MATLAB to produce tables and various graphs for analyzing the results. The VH reports results separately for each of the three sustained vowels, while it combines the results from the monologue and reading tasks.

3.2.1 Sustained vowel /a/

For this analysis, the average of the three different text files generated by the VH is used, as it provides a more reliable and robust representation of the data. This approach helps to smooth out any anomalies or errors that may occur in a single recording, offering a more accurate assessment of the overall vocal performance.

The extracted parameters are summarized in the following tables, providing a clear overview of the key vocal metrics measured during the study.

Table 3.36: Values of Parameters VH Vowel - T0.

T0					
ID	**F0 (Hz)**	**Jitt (%)**	**Shim (%)**	**$CPPS_{Median}$ (dB)**	**$CPPS_{std}$ (dB)**
2	211	0,23	1,1	18,1	1,05
4	110	0,45	3,7	16,7	1,48
7	99	0,42	5,1	15,3	1,15
9	174	1,17	4,7	17,0	1,85
10	119	0,33	2,4	18,7	1,65
12	165	0,64	1,9	16,9	1,72
14	156	1,69	11,6	12,0	3,18
16	269	0,26	1,9	15,1	1,16
17	108	0,64	5,2	17,2	1,66
21	234	0,41	2,4	15,4	1,09
1	136	0,78	2,6	16,7	2,76
3	154	1,61	6,0	10,8	1,94
5	152	0,65	2,5	16,8	1,57
6	147	1,28	5,8	17,6	3,20
8	157	1,05	3,4	13,5	1,90
11	92	1,38	8,5	13,0	1,79
13	163	0,70	2,1	15,4	1,20
15	172	0,51	1,5	15,6	1,38
19	192	0,47	2,2	17,2	1,47
20	121	2,72	6,7	15,5	3,91

Table 3.37: Values of Parameters VH vowel - T1. Light blue for negative effects of rehabilitation.

T1					
ID	**F0 (Hz)**	**Jitt (%)**	**Shim (%)**	**$CPPS_{Median}$ (dB)**	**$CPPS_{std}$ (dB)**
2	213	0,18	1,09	18,47	0,86
4	144	0,22	1,86	18,60	1,36
7	180	0,25	11,55	17,20	1,38
9	139	1,18	10,00	17,43	0,88
10	156	0,21	1,74	18,60	1,50
12	229	0,32	8,04	18,33	1,95
14	265	1,12	6,37	13,80	1,90
16	252	1,09	3,28	15,80	1,79
17	103	0,46	4,36	18,77	1,82
21	210	0,40	2,72	15,97	1,16
1	151	0,59	3,13	17,40	1,81
3	120	0,55	3,56	14,80	1,87
5	164	0,38	1,61	17,83	1,74
6	127	0,56	4,02	18,67	1,27
8	203	0,20	1,57	17,77	0,90
11	103	0,99	5,13	13,43	1,55
13	188	0,48	2,07	15,53	1,02
15	215	0,26	1,23	17,20	0,96
19	184	0,82	2,61	15,53	1,45
20	145	1,01	3,51	18,40	2,60

The Tables display all the metrics used to evaluate the sustained vowel at T0 and T1, highlighting the rehabilitation progress for all patients. At the top of the tables, patients treated with tele-rehabilitation are listed, and their ID rows are shaded in gray, at the bottom, the patients who underwent in-clinic rehabilitation are shown. The expected trends for each parameter between T0 and T1 are summarized as follows:

- **Jitter (%):** T1 values lower than T0.
- **Shimmer (%):** T1 values lower than T0.
- **CPPS median (dB):** T1 values higher than T0.
- **CPPS std (dB):** T1 values lower than T0.

The F0 parameter is reported for both T0 and T1, but it is not considered a key indicator of the technique's benefits, nevertheless, its value is included as it is essential for the accurate assessment of other parameters. In the Tables, values that diverge from the expected theoretical trend are highlighted with light blue shading, in contrast yellow shading indicates values that are the same between T0 and T1. These unexpected trends can stem from various factors, such as the fact that recordings at T0 are taken in a smaller room compared to T1, where recordings occur in a larger room, or they may fail to perform their exercises properly, or simply be having an off day. This color-coding enables a clear evaluation of the rehabilitation success rate and facilitates a comparison between the two different rehabilitation techniques, allowing for an insightful analysis of their effectiveness.

Table 3.38: Percentage of tele-rehabilitation effects VH - vowel

Measure	**Negative**	**Positive**	**None**
Jitter (%)	20%	80%	-
Shimmer (%)	50%	50%	-
$\text{CPPS}_{\text{Median}}$ (dB)	10%	90%	-
CPPS_{std} (dB)	20%	80%	-

Table 3.39: Percentage of in-clinc rehabilitation effects VH - vowel

Measure	Negative	Positive	None
Jitter (%)	10%	90%	-
Shimmer (%)	20%	80%	-
$CPPS_{Median}$ (dB)	10%	90%	-
$CPPS_{std}$ (dB)	-	100%	-

By comparing the results of tele-rehabilitation and in-clinic rehabilitation, it's possible to draw the following conclusions:

- **Jitter:** Both methods show comparable positive outcomes, with 90% of patients improving through in-clinic rehabilitation and 80% through tele-rehabilitation, indicating that both approaches are effective in reducing jitter.

- **Shimmer:** In-clinic rehabilitation proves to be more effective, with 80% of patients showing positive effects compared to only 50% for tele-rehabilitation, suggesting a greater impact on vocal stability in the clinic setting.

- **CPPS Median:** Both rehabilitation methods yield similar results, with 90% of patients showing improvement, reflecting a consistent recovery of vocal quality across both approaches.

- **CPPS Standard Deviation:** In-clinic rehabilitation stands out with a 100% success rate, while tele-rehabilitation shows an 80% improvement, indicating that in-person sessions may offer better control over voice quality consistency.

This comparison highlights that while both methods are generally effective, in-clinic rehabilitation tends to offer slightly better results, especially in parameters like shimmer and CPPS variability. However, tele-rehabilitation still provides significant improvements, making it a viable option for many patients. To provide a visual understanding of the comparison, a graph illustrating the differences between tele-rehabilitation and in-clinic rehabilitation is presented, focusing on the parameters Jitter, Shimmer, and CPPS median along with their respective standard deviations. This graph helps in highlighting the key distinctions and similarities between the two rehabilitation methods across the selected measures.

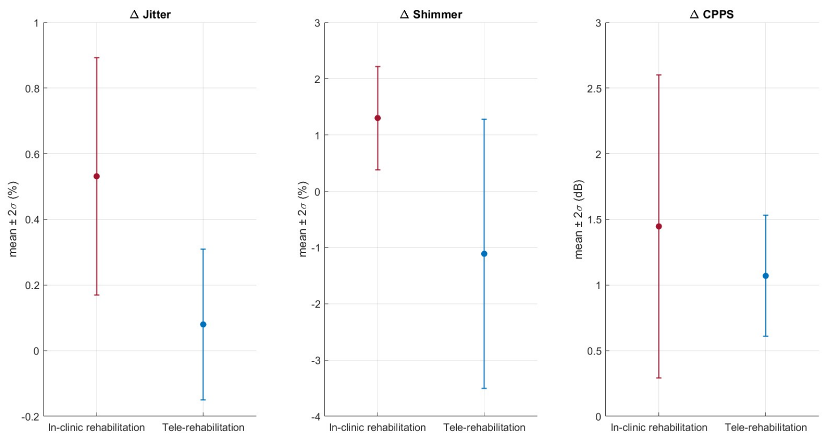


Figure 3.12: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for vowel with VH.

The Figures compare in-clinic and tele-rehabilitation groups by calculating the average difference between T1 and T0 for each parameter and their standard deviations. For Jitter and Shimmer, where a reduction at T1 is expected, the difference is calculated with a negative sign, instead for CPPS, which is expected to increase, a positive sign is applied. In the plots, the circles represent the mean values, while the error bars show twice the standard deviation divided by the square root of the 10 patients in each group (as explained in 3.1.1), this provides insight into the variability and uncertainty around the means. The overlap of confidence intervals between the two rehabilitation methods indicates that they are comparable across all parameters. Therefore, tele-rehabilitation performs similarly to in-clinic rehabilitation for the vowel study. This conclusion is further supported by the data presented in the accompanying table, reinforcing the graphical analysis.

Table 3.40: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for vowel with VH.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
Jitter (%)	0,08	0,36	0,53	0,57
Shimmer (%)	-1,1	3,8	1,3	1,5
$\text{CPPS}_{\text{Median}}$ (dB)	1,1	0,7	1,4	1,8

The Table highlights notable differences between in-clinic and tele-rehabilitation outcomes, particularly in certain parameters, for instance, Shimmer shows the most significant discrepancy, with an average difference between T1 and T0 of 1.3% in in-clinic rehabilitation, while tele-rehabilitation exhibits a negative value of -1.1%. This contrast is especially striking. Jitter and CPPS median also display differences, with both parameters having higher values in in-clinic rehabilitation.

When examining the standard deviation of these parameters, Shimmer again stands out, in tele-rehabilitation, its standard deviation is 3.8%, more than double the 1.5% observed in in-clinic rehabilitation, which is concerning, as a higher standard deviation indicates greater data variability. Jitter, on the other hand, shows slightly lower values in tele-rehabilitation compared to in-clinic, while CPPS median has a notably higher standard deviation in in-clinic rehabilitation.

Despite the differences highlighted here, the Figure 3.1 shows an overlap in the reported bands. This indicates that, even if only partially, the two rehabilitation techniques are still comparable.

3.2.2 Monologue and reading

In VH, as explained in 2.5, two different text files are produced, and in this section, the 'Parameters Tab' is considered, where it is not possible to distinguish between monologue and reading since VH records both together, reporting only the point-by-point values that include both. The parameters of interest are SPL, F0, and CPPS, along with their statistics; however, in this case, the focus is on the mean values. There are some issues with these files because some of them are not recorded correctly, resulting in missing or unreported values, so it is not possible to consider the following individuals:

- **Removed at T0:** Patients 1, 2, 3, and 6.

- **Removed at T1:** Patients 1, 2, and 17.

Therefore, to conduct an evaluation between T0 and T1, all of these patients need to be excluded, resulting in a reduction of the dataset to 15 patients instead of 20. Another important consideration is that, as previously explained, VH reports SPL at 22 cm, but a conversion to 30 cm is now required because this is the standard value used in the literature, and the values reported from in-air microphones are measured at this distance, consequently, the results reported below are adjusted for 30 cm.

Table 3.41: Values of VH monologue and reading parameters - T0.

T0			
ID	SPL (dB)	F0 (Hz)	CPPS (dB)
2	-	-	-
4	76,9	108	12,6
7	77,0	108	13,3
9	72,3	128	11,6
10	74,2	95	14,5
12	68,0	145	14,6
14	69,8	117	9,0
16	71,9	198	13,5
17	-	-	-
21	73,5	158	14,4
1	-	-	-
3	-	-	-
5	72,3	142	15,2
6	-	-	-
8	74,5	161	13,1
11	66,5	109	11,1
13	69,4	152	13,6
15	70,0	154	14,5
19	73,1	126	13,4
20	75,8	104	12,2

Table 3.42: Values of VH monologue and reading parameters - T1.

T1			
ID	SPL (dB)	F0 (Hz)	CPPS (dB)
2	-	-	-
4	75,0	101	12,5
7	78,7	107	13,6
9	72,7	167	12,6
10	78,9	107	13,3
12	82,1	172	15,1
14	69,4	138	10,3
16	75,2	197	13,1
17	-	-	-
21	76,4	186	14,7
1	-	-	-
3	-	-	-
5	70,5	125	15,0
6	-	-	-
8	76,4	184	14,5
11	71,2	123	11,1
13	70,4	152	13,5
15	77,8	176	16,0
19	90,2	145	13,6
20	76,7	115	12,8

While F0 values are included in the Tables, they are excluded from the analysis of T1 and T0, the values marked in light blue diverge from the expected theoretical trend, indicating that the rehabilitation may not have had the desired positive impact. Notably, 8 patients underwent tele-rehabilitation, while 7 received in-clinic treatment. It is possible to determine the percentage of effect achieved by the rehabilitation

Table 3.43: Percentage of tele-rehabilitation effects VH - monologue and reading.

Measure	Negative	Positive	None
SPL (dB)	25%	75%	-
CPPS (dB)	37,5%	62,5%	-

Table 3.44: Percentage of in-clinc rehabilitation effects VH - monologue and reading.

Measure	Negative	Positive	None
SPL (dB)	14,3%	85,7%	-
CPPS (dB)	28,6%	71,4%	-

By comparing the results of tele-rehabilitation and in-clinic rehabilitation, we can draw the following conclusions:

- **SPL (dB):** Tele-rehabilitation demonstrates a significant positive effect, with 75% of patients reporting improvements, while in-clinic rehabilitation shows an even higher rate of 85.7% positive outcomes, suggesting that both methods effectively enhance sound pressure levels.

- **CPPS (dB):** Both rehabilitation approaches yield positive results, but tele-rehabilitation has a lower percentage of patients showing improvement (62.5%) compared to in-clinic rehabilitation (71.4%). This indicates a potentially greater effectiveness of in-clinic sessions in improving vocal quality.

While tele-rehabilitation presents a substantial proportion of positive effects, in-clinic rehabilitation outperforms it in both measures. This suggests that while tele-rehabilitation is beneficial, in-clinic sessions may provide a more comprehensive improvement in vocal measures. It is crucial to acknowledge that the dataset comprises only 15 patients, which, while sufficient for preliminary analysis, restricts the generalizability of the findings.

Therefore, following this initial analysis using the "Parameters Tab," which reports the mean values obtained from both monologue and reading, it is essential to conduct another study to investigate in-clinic and tele-rehabilitation in more detail. As discussed in Section 2.5, VH generated another file titled 'Parameters 46 ms', which contains frame-by-frame values for F0 and SPL, including both reading and monologue sections. By considering the period of silence between them, it is possible to separate the file into the respective reading and monologue portions. While Section 3.3.2 presents the frame-by-frame trends, this file is used here to calculate and evaluate the mean SPL values, because in this work VH provides only the mean SPL values for both the reading and monologue. This analysis further illuminates the effectiveness of the rehabilitation methods by providing a clearer understanding of vocal performance trends over time.

Reading

For the reading task, the first part of the 'Parameters 46ms' file is used, after removing the zeros, which represent silent frames, the values for each voice frame are extracted. By calculating the mean of these values, it is possible to obtain an estimation of the SPL and F0 means for VH during the reading.

Table 3.45: Values of VH reading parameters - T0.

T0		
ID	**F0 (Hz)**	**SPL (dB)**
2	-	-
4	110	77,5
7	112	77,3
9	140	72,3
10	98	74,7
12	159	67,6
14	189	71,3
16	173	72,4
17	-	-
21	154	73,7
1	-	-
3	-	-
5	143	72,7
6	-	-
8	164	74,7
11	109	66,9
13	177	70,1
15	163	69,4
19	107	73,3
20	104	75,8

Table 3.46: Values of VH reading parameters - T1.

T1		
ID	**F0 (Hz)**	**SPL (dB)**
2	-	-
4	102	75,0
7	107	79,0
9	192	72,9
10	110	79,5
12	179	79,7
14	172	69,3
16	208	75,6
17	-	-
21	196	76,8
1	-	-
3	-	-
5	127	71,4
6	-	-
8	203	76,9
11	127	71,3
13	150	70,9
15	183	78,1
19	164	89,9
20	118	77,0

With these values, it is possible to compare in-clinic and tele-rehabilitation treatments, noting that in the Tables, the first gray rows correspond to patients treated with tele-rehabilitation, while the latter rows represent in-clinic patients.

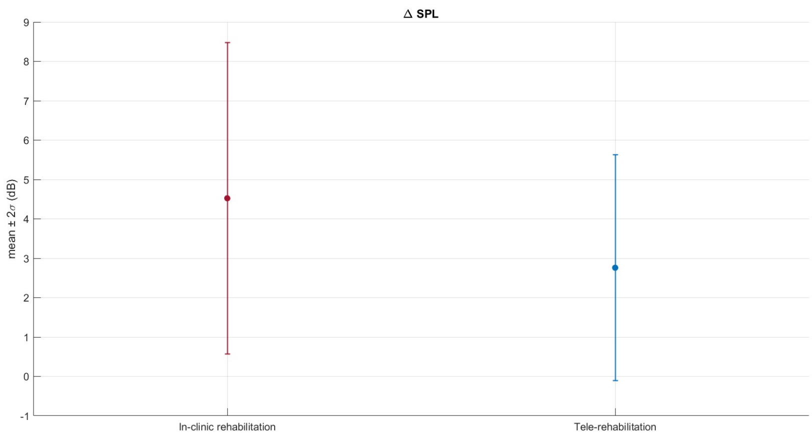


Figure 3.13: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for reading with VH.

The Figure illustrates the comparison between in-clinic and tele-rehabilitation groups by calculating the average change from T0 to T1 for each parameter, along with their corresponding standard deviations. In the charts, mean values are shown as circles, while the error bars depict two times the standard deviation, adjusted by dividing by the square root of the number of patients in each group: 8 for tele-rehabilitation and 7 for in-clinic (as detailed in 3.1.1). This highlights both the variation and uncertainty of the mean values. The fact that the confidence intervals overlap between the two rehabilitation approaches indicates that they perform similarly across all parameters. This finding is also backed by the table data, providing additional confirmation to the visual results.

Table 3.47: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for reading with VH.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
SPL (dB)	2,8	4,5	4,5	6,3

The Table highlights the differences between the various rehabilitation methods regarding the SPL parameter. In-clinic rehabilitation demonstrates a higher mean difference value between T1 and T0 compared to tele-rehabilitation; however, it also exhibits a greater mean standard deviation, indicating significant variability within this patient group.

Monologue

For the monologue task, the second part of the 'Parameters 46ms' file is used, after removing the zeros, which represent silent frames, the values for each voice frame are extracted. By calculating the mean of these values, it is possible to obtain an estimation of the SPL and F0 means for VH during the reading.

Table 3.48: Values of VH monologue parameters - T0.

T0		
ID	F0 (Hz)	SPL (dB)
2	-	-
4	106	77,0
7	104	77,2
9	142	73,8
10	91	74,1
12	132	68,9
14	127	70,3
16	224	71,8
17	-	-
21	161	73,7
1	-	-
3	-	-
5	141	72,4
6	-	-
8	157	74,8
11	109	66,4
13	127	69,2
15	150	71,2
19	145	73,5
20	103	77,2

Table 3.49: Values of VH monologue parameters - T1.

T1		
ID	F0 (Hz)	SPL (dB)
2	-	-
4	102	75,4
7	108	78,7
9	143	73,0
10	103	78,7
12	165	85,0
14	104	70,2
16	176	75,2
17	-	-
21	168	76,3
1	-	-
3	-	-
5	122	70,5
6	-	-
8	167	76,2
11	116	71,6
13	154	70,3
15	169	78,1
19	116	90,8
20	108	77,2

Using these values, a comparative analysis can be conducted between the treatments administered in-clinic and those provided through tele-rehabilitation. It is important to note that in the Tables presented, the initial rows shaded in gray pertain to patients who underwent tele-rehabilitation. In contrast, the subsequent rows depict the data for patients who received their treatment in a traditional clinic setting. This distinction allows for a clearer understanding of the differences and potential outcomes associated with each rehabilitation method.

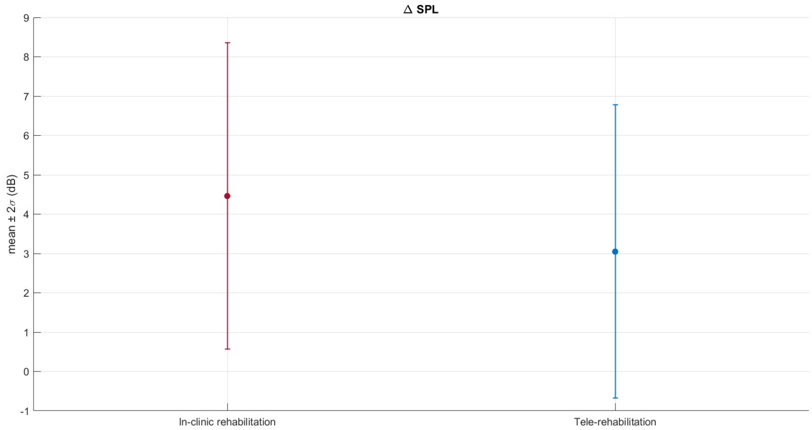


Figure 3.14: Comparison of tele-rehabilitation and in-clinic rehabilitation methods for monologue with VH.

The Figure provide a visual comparison of the in-clinic and tele-rehabilitation groups by analyzing the average differences between T0 and T1 for each parameter, along with their associated standard deviations. In these plots, the mean values are represented by circles, while the error bars illustrate twice the standard deviation, adjusted for the sample size in each group 8 patients for tele-rehabilitation and 7 patients for in-clinic treatment (as explained in 3.1.1). This effectively demonstrates the variability and uncertainty surrounding the mean values. The overlapping confidence intervals for both rehabilitation methods suggest that they yield comparable results across all assessed parameters. Additionally, the data presented in the accompanying table further reinforces these conclusions, corroborating the insights gleaned from the graphical representation.

Table 3.50: Comparison of average differences and standard deviations between T0 and T1 for Tele-rehabilitation and In-clinic rehabilitation for monologue with VH.

	Tele-rehabilitation		**In-clinic rehabilitation**	
	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$	$\overline{x_{T1-T0}}$	$\overline{\sigma_{T1-T0}}$
SPL (dB)	3,0	5,9	4,5	6,9

The Table highlights the differences between the various rehabilitation methods regarding the SPL parameter. In-clinic rehabilitation demonstrates a higher mean difference value between T1 and T0 compared to tele-rehabilitation; however, it also exhibits a greater mean standard deviation, indicating significant variability within this patient group. Overall, while in-clinic rehabilitation appears to yield a greater average improvement in SPL, the accompanying variability may warrant further investigation to understand the underlying factors contributing to this difference. In contrast, tele-rehabilitation demonstrates more consistent outcomes, suggesting it may provide a stable alternative for patients requiring vocal rehabilitation.

3.3 Comparison between in-air Microphone and Vocal Holter device

Following the initial analysis, which provides a comprehensive overview and detailed explanation of all the parameters related to the two acquisition systems, this section aims to conduct a comparison between the in-air microphone and the vocal Holter device. This comparison highlights the differences and similarities in terms of performance, accuracy, and applicability of both systems, thereby offering a clearer understanding of their respective functionalities and the contexts in which each device proves most effective. As previously explained, the Vocal Holter uses the same MATLAB scripts with an HNR threshold set at 0 dB, therefore the comparison is made with the results from the in-air microphone, without considering the -10 dB and -20 dB thresholds. Although both VH and the in-air microphone utilize the same MATLAB scripts, they acquire voice signals differently, the VH, being a contact microphone, captures the vibrations of the vocal folds, while the in-air microphone records the sound waves in the surrounding environment.

3.3.1 Sustained vowel /a/

For the sustained vowel /a/, the parameters under investigation are F0, Jitter percentage, Shimmer percentage, CPPS median, and CPPS standard deviation. These parameters are derived from the Vocal Holter data, although other parameters can be extracted using a microphone, only these specific metrics are used for comparison purposes. First, to understand the differences between the VH and the in-air microphone, a dispersion graph is reported.

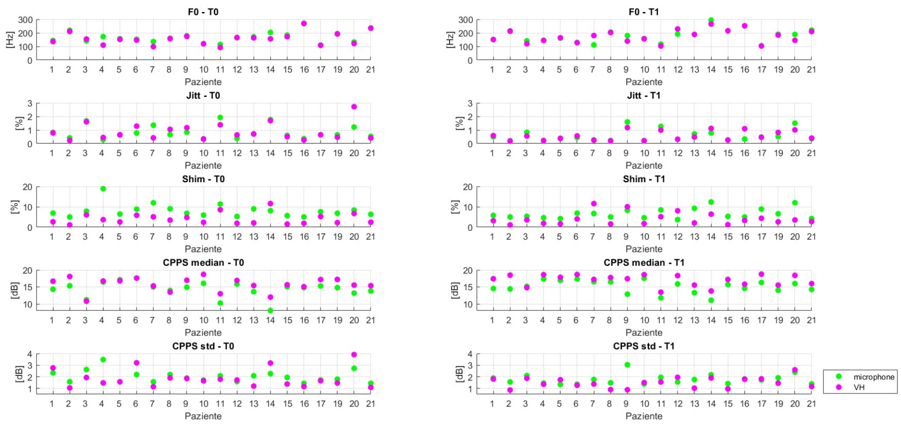


Figure 3.15: Distribution of parameter values extracted separately for T0 and T1 using the Vocal Holter and the in-air microphone for the sustained vowel.

To gain a clearer understanding of the differences between VH and the in-air microphone as illustrated in the Figures, a detailed evaluation is presented in the following Tables where the Vocal Holter device is defined as 'VH' while 'MIC' refers to the in-air microphone.

Table 3.51: Comparison of parameters between VH and in-air microphone, showing microphone values and the delta from VH values for each patient at T0 - vowel.

	F0 (Hz)		**Jitter (%)**		**Shimmer (%)**		**CPPS$_{\mathbf{Median}}$ (dB)**		**CPPS$_{\mathbf{std}}$ (dB)**	
ID	**MIC**	Δ_{VH-MIC}	**MIC**	Δ_{VH-MIC}	**MIC**	Δ_{VH-MIC}	**MIC**	Δ_{VH-MIC}	**MIC**	Δ_{VH-MIC}
1	143	-7	0,82	-0,04	6,9	-4,3	14,3	2,3	2,33	0,43
2	219	-8	0,42	-0,19	5,0	-3,9	15,3	2,7	1,58	-0,53
3	141	13	1,67	-0,06	7,8	-1,8	11,2	-0,4	2,62	-0,69
4	172	-62	0,29	0,16	18,9	-15,2	16,4	0,3	3,48	-2,00
5	157	-5	0,63	0,02	6,4	-3,9	17,1	-0,3	1,57	0,00
6	152	-6	0,78	0,51	8,8	-3,0	17,6	0,0	2,19	1,01
7	136	-37	1,35	-0,93	12,0	-6,9	15,0	0,3	1,56	-0,42
8	160	-3	0,67	0,38	9,0	-5,6	13,9	-0,5	2,19	-0,29
9	179	-4	0,84	0,33	6,8	-2,0	14,9	2,0	1,90	-0,05
10	121	-2	0,36	-0,03	5,9	-3,5	16,0	2,7	1,72	-0,07
11	114	-22	1,92	-0,54	11,3	-2,8	10,2	2,8	2,09	-0,30
12	167	-2	0,38	0,25	5,3	-3,4	15,8	1,1	1,59	0,14
13	171	-7	0,72	-0,02	8,9	-6,9	13,6	1,8	2,09	-0,89
14	203	-47	1,77	-0,08	8,1	3,5	8,0	4,0	2,27	0,91
15	183	-11	0,60	-0,09	5,6	-4,1	15,1	0,6	1,95	-0,57
16	271	-2	0,37	-0,11	5,0	-3,1	14,8	0,3	1,44	-0,27
17	110	-2	0,65	-0,01	7,5	-2,4	15,3	1,9	1,72	-0,06
19	194	-2	0,64	-0,17	6,8	-4,6	14,8	2,4	1,79	-0,33
20	132	-11	1,22	1,50	8,3	-1,6	13,2	2,4	2,72	1,19
21	237	-3	0,53	-0,12	6,3	-3,9	13,8	1,6	1,44	-0,35

Table 3.52: Comparison of parameters between VH and in-air microphone, showing microphone values and the delta from VH values for each patient at T1 - vowel.

	F0 (Hz)		**Jitter (%)**		**Shimmer (%)**		**$CPPS_{Median}$ (dB)**		**$CPPS_{std}$ (dB)**	
ID	**MIC**	Δ_{VH-MIC}	**MIC**	Δ_{VH-MIC}	**MIC**	Δ_{VH-MIC}	**MIC**	Δ_{VH-MIC}	**MIC**	Δ_{VH-MIC}
1	150	1	0,50	0,09	5,7	-2,6	14,5	2,9	1,89	-0,09
2	215	-2	0,21	-0,03	5,0	-3,9	14,4	4,1	1,56	-0,69
3	141	-21	0,83	-0,28	5,3	-1,8	15,2	-0,4	2,10	-0,23
4	146	-2	0,23	-0,02	4,6	-2,7	17,3	1,3	1,46	-0,10
5	161	2	0,38	0,00	4,2	-2,6	16,9	0,9	1,35	0,40
6	129	-2	0,44	0,12	6,9	-2,8	17,3	1,3	1,36	-0,08
7	112	69	0,28	-0,03	6,7	4,9	16,5	0,7	1,76	-0,39
8	205	-3	0,23	-0,03	5,0	-3,4	16,5	1,3	1,48	-0,58
9	179	-40	1,59	-0,41	8,3	1,7	12,9	4,6	3,03	-2,15
10	158	-2	0,20	0,02	4,6	-2,9	17,6	1,0	1,38	0,12
11	117	-13	1,26	-0,27	8,4	-3,3	11,8	1,6	1,94	-0,39
12	191	37	0,32	0,00	3,7	4,4	15,9	2,4	1,55	0,40
13	189	-1	0,71	-0,23	9,2	-7,1	13,3	2,2	1,76	-0,74
14	294	-29	0,79	0,33	12,3	-6,0	11,1	2,7	2,17	-0,27
15	216	0	0,27	-0,02	5,3	-4,0	15,7	1,5	1,41	-0,45
16	252	0	0,34	0,76	5,0	-1,7	14,6	1,2	1,82	-0,04
17	105	-2	0,49	-0,02	8,9	-4,5	16,3	2,5	1,72	0,10
19	190	-6	0,51	0,31	6,6	-3,9	14,0	1,5	1,91	-0,46
20	188	-43	1,50	-0,49	12,0	-8,4	16,0	2,4	2,40	0,20
21	220	-10	0,35	0,05	4,2	-1,5	14,3	1,7	1,39	-0,22

In the Table, the values obtained from the in-air microphone are reported, along with the difference between the VH and the in-air microphone at T0 and T1, this highlights the sign of the difference and helps to understand how these acquisition systems compare, in detail:

- **F0:** At T0, all but one value decrease in VH compared to the in-air microphone. At T1, the majority of parameters also show a decrease with VH; however, three patients exhibit an increase, and two remain unchanged.
- **Jitter:** At T0, approximately 13 out of 20 values decrease in VH relative to the in-air microphone. By T1, 11 patients show a decrease with VH, while two maintain their values.
- **Shimmer:** All values decrease in VH at T0 except for one. This trend continues at T1, where three values increase, indicating some variability.
- **CPPS Median:** At T0, all values but three increase in VH compared to the in-air microphone. At T1, a similar pattern is observed, with all values increasing except for one.
- **CPPS std:** At T0, about 13 out of 20 values decrease in VH when compared to the in-air microphone. This trend becomes more pronounced at T1, where 15 out of 20 values show a decrease.

By analyzing the mean values and the mean differences between the results obtained with the VH and the in-air microphone, a clearer understanding of the overall trends can be achieved. This analysis facilitates the identification of patterns and variations that may not be immediately apparent. Furthermore, to gain a more comprehensive insight, the standard deviation values have also been evaluated, providing additional information about the variability of the collected data. The tables presented below display the average values for 20 patients and their corresponding standard deviations at T0 and T1, allowing for a detailed comparison between the two acquisition systems.

Table 3.53: Average values and deltas between MIC and VH for T0 vowel.

Average values for 20 patients		
Parameters	**MIC**	Δ_{VH-MIC}
F0 (Hz)	168	-12
Jitter (%)	0,83	0,04
Shimmer (%)	8,0	-4,0
CPPS $_{\textbf{Median}}$ (dB)	14,3	1,4
CPPS $_{\textbf{std}}$ (dB)	2,01	-0,16

Table 3.54: Average values and deltas between MIC and VH for T1 vowel.

Average values for 20 patients		
Parameters	**MIC**	Δ_{VH-MIC}
F0 (Hz)	178	-3
Jitter (%)	0,57	-0,01
Shimmer (%)	6,6	-2,6
CPPS $_{\textbf{Median}}$ (dB)	15,1	1,9
CPPS $_{\textbf{std}}$ (dB)	1,77	-0,28

Table 3.55: Standard deviation values and deltas between MIC and VH for T0 vowel.

Standard deviation values for 20 patients		
Parameters	**MIC**	Δ_{VH-MIC}
F0 (Hz)	42	4
Jitter (%)	0,49	0,13
Shimmer (%)	3,2	-0,5
CPPS $_{\textbf{Median}}$ (dB)	2,3	-0,2
CPPS $_{\textbf{std}}$ (dB)	0,51	0,28

Table 3.56: Standard deviation values and deltas between MIC and VH for T1 vowel.

Standard deviation values for 20 patients		
Parameters	**MIC Value**	Δ_{VH-MIC}
F0 (Hz)	48	-1
Jitter (%)	0,43	-0,08
Shimmer (%)	2,5	0,4
CPPS $_{\textbf{Median}}$ (dB)	1,8	-0,2
CPPS $_{\textbf{std}}$ (dB)	0,42	0,03

The analysis of average values reveals a general decrease in the VH readings, while the CPPS median shows an increase in both T0 and T1 cases. The differences observed between these two acquisition systems can be attributed to the distinct methods of voice signal capture. For instance, the in-air microphone is more susceptible to background noise and ambient conditions, as well as potential issues related to microphone positioning or calibration. Notably, the parameter showing the greatest difference between the two microphones is Shimmer, it is particularly sensitive to background noise and other sound sources, which results in lower values for VH. In contrast, the in-air microphone captures all sounds in the patient's environment, leading to higher Shimmer values, since Shimmer represents a peak-to-peak evaluation, it is significantly influenced by these external factors. The median values of CPPS show an increase for the VH microphone compared to the in-air microphone, this difference can be attributed to the distinct bandwidths of the two microphones. The VH microphone is in contact with the neck, which introduces a low-pass filter effect, attenuating higher frequencies before they are recorded, in contrast the in-air microphone captures a broader range of frequencies. Additionally, background noise can significantly impact the accuracy of CPPS evaluations. The presence of ambient noise may distort the readings from the in-air microphone, while the VH microphone's proximity to the neck may help mitigate some of this interference, leading to a higher recorded value for CPPS.

Regarding the standard deviation values, it is important to emphasize that these represent the variability of the measurements within the data, lower values indicate greater consistency in the results, while higher values suggest more significant variability. Looking at the data, it's possible to observe that some standard deviations, such as that of the Shimmer parameter, show notable differences between the two microphones. The greater variability in measurements from the in-air microphone may be attributed to its sensitivity to external factors, such as background noise, in contrast, the VH microphone, with a lower standard deviation, provides more stable and reliable results. This information is clinically relevant, as a lower standard deviation for the VH microphone suggests that this tool may yield more repeatable measurements, which are essential for diagnosis and patient monitoring. Ultimately, the choice of microphone not only affects the average values but also the variability of the results, with significant implications for voice assessments.

3.3.2 Monologue and reading

In VH, as explained in 3.2.2, two different text files are generated,the 'Parameters Tab' report includes point-by-point values for both monologue and reading, making it impossible to distinguish between them, in contrast the 'Parameters 46 ms' file allows for separate analysis of monologue and reading. This file provides frame-by-frame values for F0 and SPL, since the VH recordings are made together but include a period of silence between the monologue and reading represented by a long sequence of zeros in this file. These zeros are highlighted to split the file into two parts, using the first section for the monologue and the second for the reading. In the case of the VH file "Parameters 46 ms," it is analyzed using a MATLAB script to exclude all zeros, allowing for a representation of only the F0 and SPL values linked to the patient's voice. Similarly, for the in-air microphone, the values obtained from frame-to-frame analysis are evaluated, noting that for males, the frequencies are defined between 75 and 300 Hz, while for females, they range from 100 to 400 Hz.

Issues arise with these files, as some are not recorded accurately, leading to missing or unreported values. Consequently, the following individuals are excluded:

- **Removed at T0:** Patients 1, 2, 3 and 6.
- **Removed at T1:** Patients 1, 2 and 17.

To proceed with the study, the dataset must be reduced from 20 to 15 patients, excluding those listed above, furthermore it is essential to note that VH reports SPL at 22 cm, that are converted to 30 cm. Additionally, since the VH contains Matlab scripts to report the different analyzed parameters, it is necessary to use an HNR threshold of 0 dB for the in-air microphone, as the VH scripts are set to this threshold.

Reading

To compare the readings from the in-air microphone and the VH, the 'Parameters 46 ms' text file is used, allowing for frame-by-frame values to be represented in histograms for both F0 and SPL. This approach enables the evaluation of differences between the two microphones. For all the histograms presented below for each patient, the gender is also indicated to provide a better understanding of the differences between male and female voices.

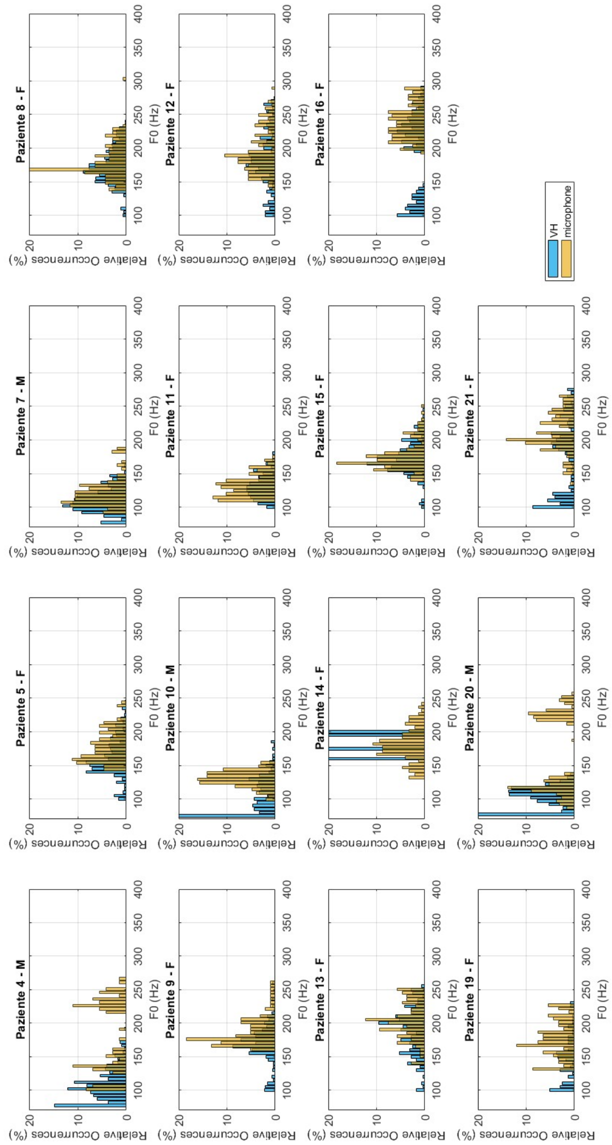


Figure 3.16: F0 Histogram Comparison for Reading T0: VH vs in-air microphone.

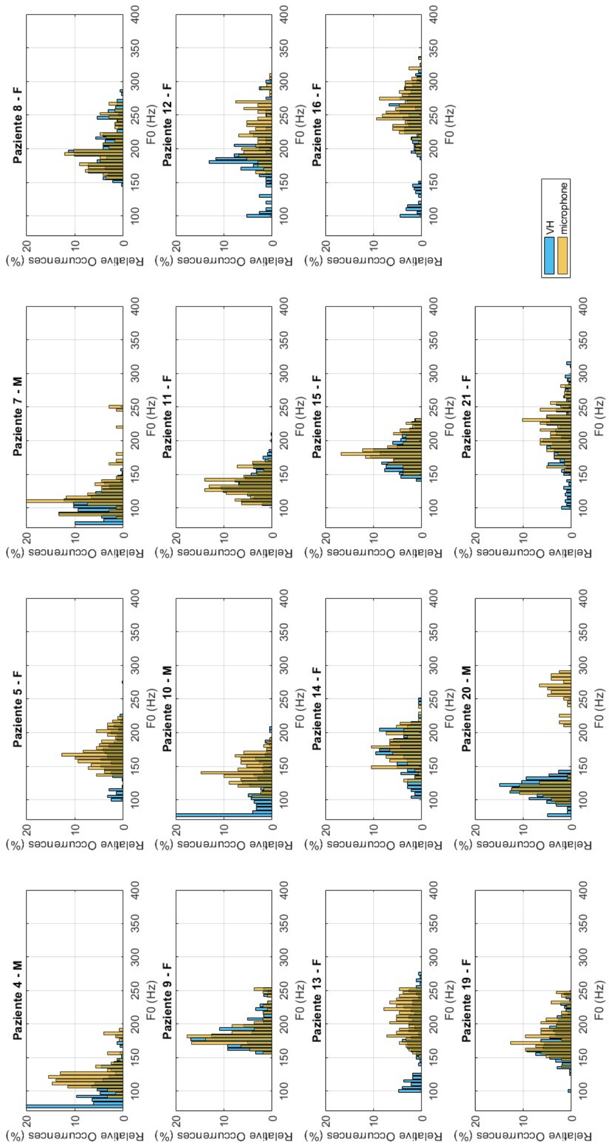


Figure 3.17: F0 Histogram Comparison for Reading T1: VH vs in-air microphone.

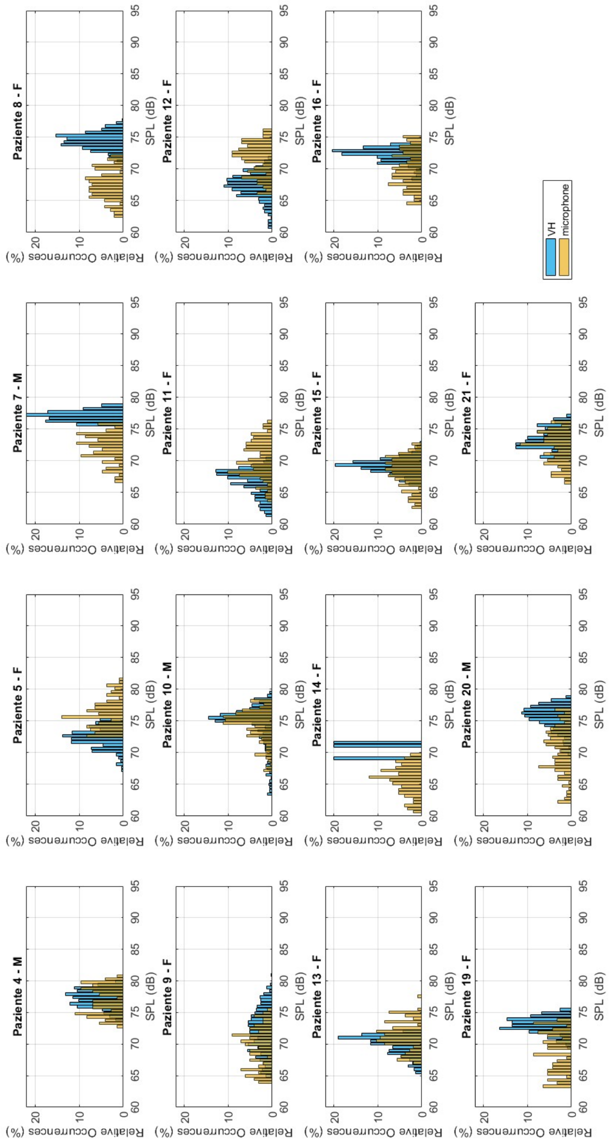


Figure 3.18: SPL Histogram Comparison for Reading T0: VH vs in-air microphone.

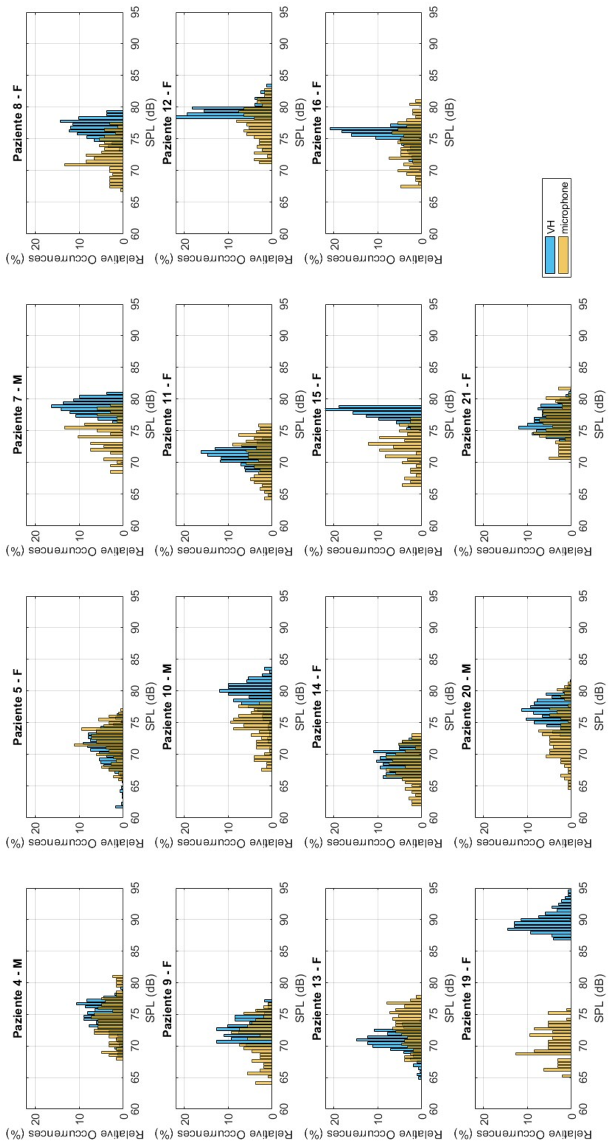


Figure 3.19: SPL Histogram Comparison for Reading T1: VH vs in-air microphone.

The F0 histograms at T0 and T1 reveal both differences and similarities between the two microphones. It is evident that for the majority of patients, the histograms are nearly overlapping, consistently showing histograms shifted to the right towards higher values for the air microphone compared to the VH microphone. In some cases, a bimodal trend can be observed for certain patients:

- **VH:** Patient 16 both for T0 and T1.
- **In-air microphone:** At T0 for patients 4 and 20, at T1 for patient 20.

The higher values and bimodal trend observed with the in-air microphone may be attributed to its ability to capture sound waves along with background noise and environmental distortions surrounding the patient. In contrast, the VH microphone detects vibrations from the vocal folds, which may result in lower values, the bimodal trend, particularly for certain patients, could indicate specific issues related to their evaluation of vocal fold vibrations. In summary, the F0 histograms comparing the VH and the in-air microphone show a significant overlap, indicating that, for the majority of patients, both devices measure the F0 consistently. This similarity in measurements suggests that, despite differences in sound acquisition methods, both microphones can be considered valid tools for assessing F0.

The SPL histograms at T0 and T1 show a slightly different trend compared to the F0 histograms. In the case of the in-air microphone, the histograms either overlap with the VH or are shifted to the right, indicating higher values. However, in the SPL histograms, it's possible to distinguish three different scenarios:

- **Overlap between the in-air microphone and VH histograms:** At T0, this is observed for patients 4, 9, 10 and 21. At T1, the overlap is evident for patients 4, 5, 9, 14 and 21 .
- **VH histogram shifted to the right compared to the in-air microphone:** At T0, this is seen in patients 7, 8 and 14. At T1, this shift is observed for patients 15 and 19.
- **In-air microphone histogram shifted to the right compared to VH:** At T0, this occurs for patients 11 and 12. At T1, the same shift is noted for patient 13.

Overall, the trend for the VH microphone is generally higher than for the in-air microphone. This difference is due to the VH microphone capturing a more focused and stable sound pressure level by directly measuring vocal cord vibrations, furthermore the lower dispersion of values in the VH recordings indicates a more stable measurement. In contrast, the in-air microphone captures sound waves along with background noise and distortions, which can impair the evaluation of SPL, additionally the range of values for the in-air microphone is greater, reflecting a wider distribution compared to the VH microphone.

Monologue

The 'Parameters 46 ms' text file is utilized to compare data from the in-air microphone and the VH, enabling the creation of histograms that display frame-by-frame values for both F0 and SPL, this method allows for an assessment of the differences between the two microphones. Additionally, the histograms for each patient include their gender, offering clearer insights into how male and female voices differ.

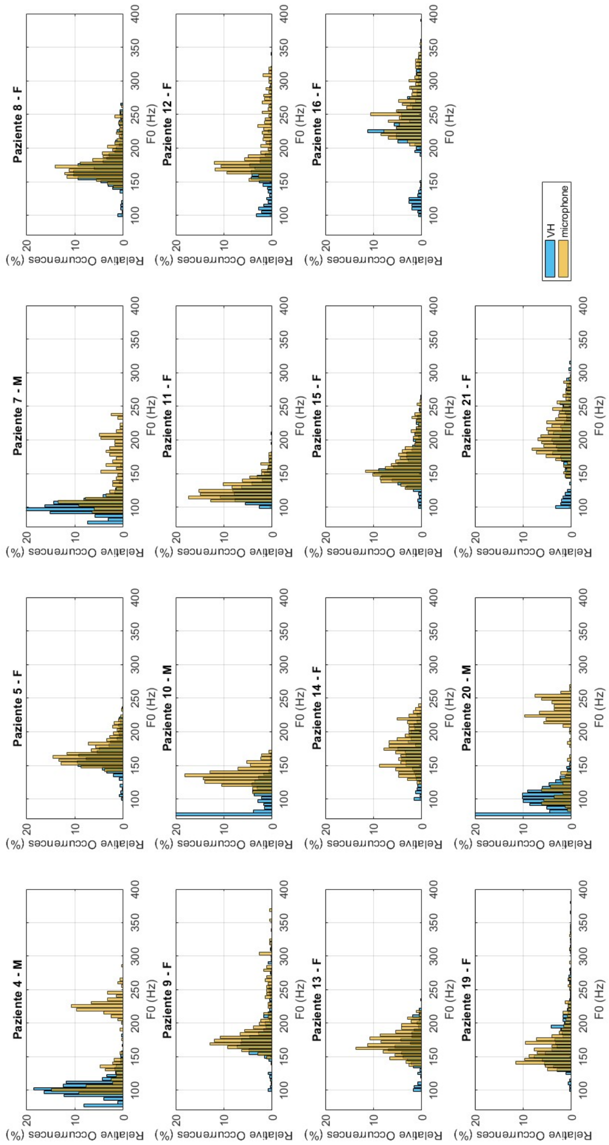


Figure 3.20: F0 Histogram Comparison for monologue T0: VH vs in-air microphone.

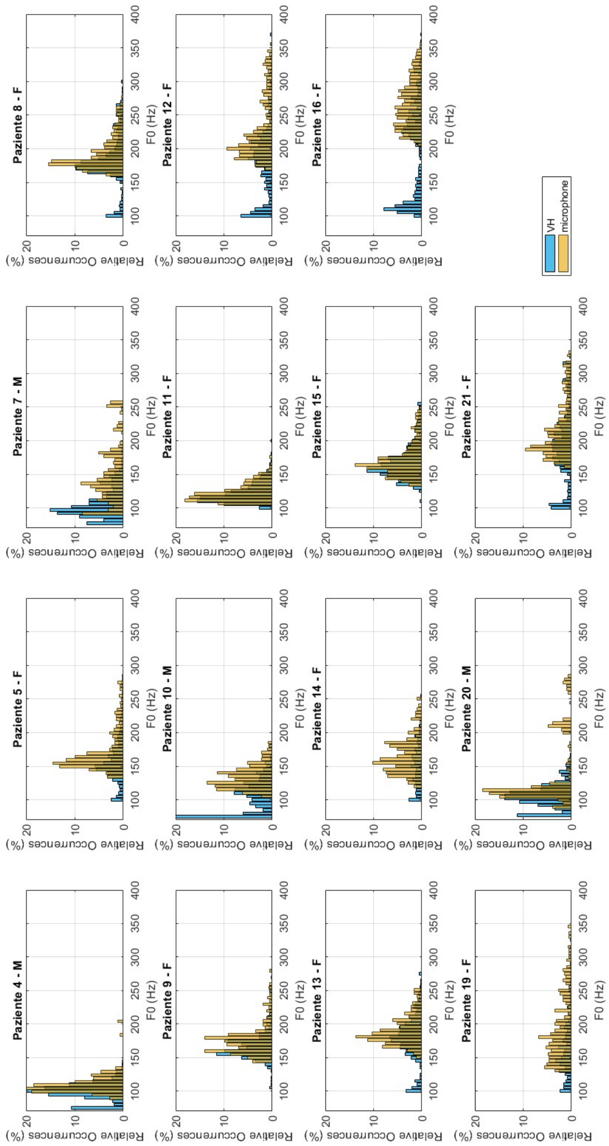


Figure 3.21: F0 Histogram Comparison for monologue T1: VH vs in-air microphone.

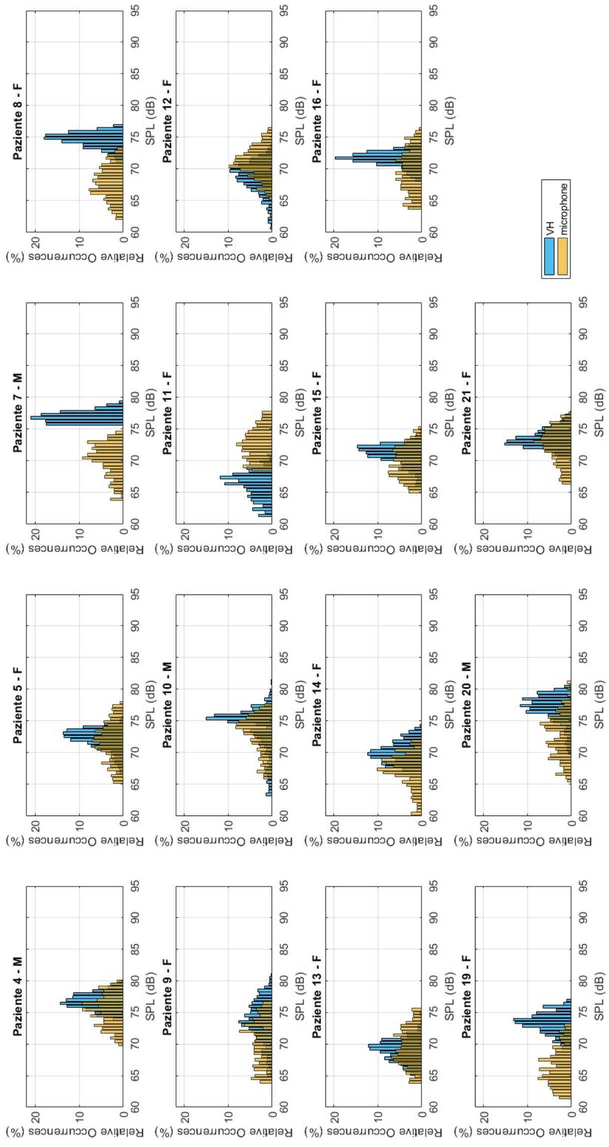


Figure 3.22: SPL Histogram Comparison for monologue T0: VH vs in-air microphone.

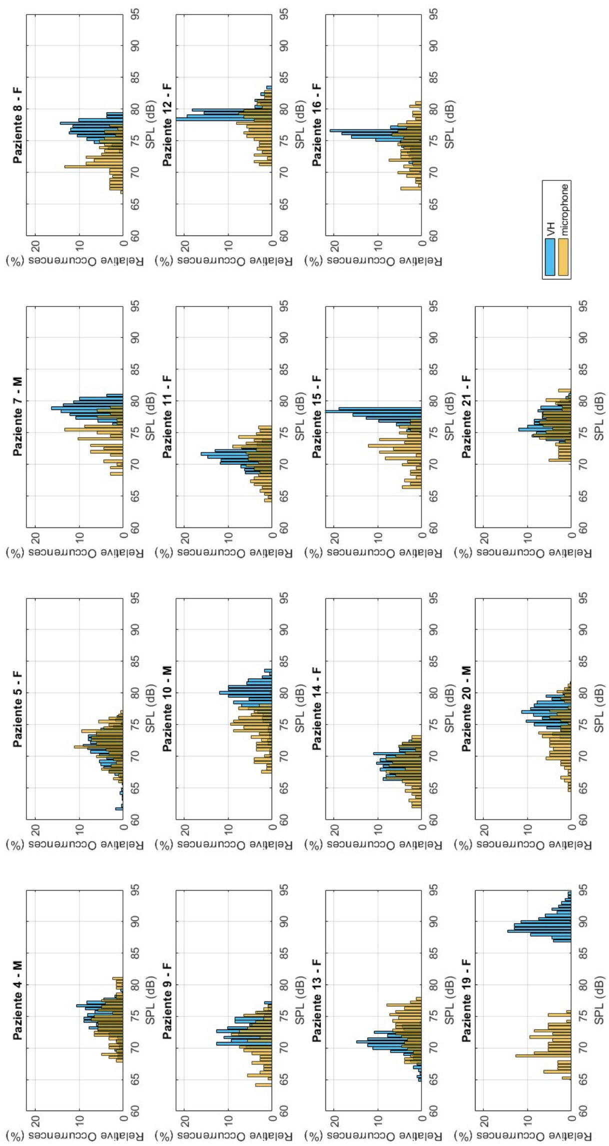


Figure 3.23: SPL Histogram Comparison for monologue T1: VH vs in-air microphone.

The F0 histograms at T0 and T1 demonstrate both differences and similarities between the two microphones. For the majority of patients, the histograms from both microphones closely align, showing only minor variations, however a consistent pattern emerges, with the histograms for the air microphone typically shifted to the right, indicating higher F0 values compared to those recorded by the VH microphone. The cases highlethed are:

- **Overlap between the in-air microphone and VH histograms:** At T0 for patients 5, 8, 9, 11, 13, 14, 15 and 19. At T1 for patients 11 and 15.
- **In-air microphone histogram shifted to the right compared to VH:** At T0 for patients 7, 10 and 12. At T1 for patients 5, 7, 10, 16, 19, 20, 21.
- **Bimodal trend:** At T0 for patients 4 and 20.

The higher values and bimodal pattern seen with the in-air microphone might be due to its ability to capture both the patient's voice and surrounding background noise or environmental distortions. In contrast, the VH microphone detects vibrations directly from the vocal folds, leading to lower readings. The bimodal trend, particularly in certain patients, could highlight specific issues in how their vocal fold vibrations are evaluated. Overall, the F0 histograms from the VH and in-air microphones show considerable overlap, indicating that both devices consistently measure F0 in most patients. This similarity suggests that, despite the differences in sound acquisition methods, both microphones are reliable tools for assessing F0.

The SPL histograms at T0 and T1 show a slightly different trend compared to the F0 histograms. It's possible to distinguish three different scenarios:

- **Overlap between the in-air microphone and VH histograms:** A satisfactory overlap is not observed in any of the histograms, either at T0 or T1.
- **VH histogram shifted to the right compared to the in-air microphone:** At T0, this is seen in patients 7, 8 ,9, 14, 19 and 20. At T1, this shift is observed for patients 7, 8, 12, 15 and 20 and also a clear separation is observed for patient 19.
- **In-air microphone histogram shifted to the right compared to VH:** At T0, this occurs for patients 11 and 12. At T1, the same shift is noted for patient 5, 13.

On the whole, the VH microphone tends to register higher SPL compared to the in-air microphone, this is primarily because the VH microphone directly measures vocal fold vibrations, resulting in a more focused and consistent SPL monologue, the narrower spread of values in the VH data also points to a more stable and reliable measurement process. On the other hand, the in-air microphone captures both the patient's voice and surrounding environmental noise, which introduces distortions and makes the evaluation of SPL less precise. As a result, the range of values for the in-air microphone is broader, indicating a wider distribution of SPL readings when compared to the more concentrated results obtained from the VH microphone.

Chapter 4

Conclusion

In this work, various studies are conducted on voice recordings of MS patients: examining the differences between in-clinic and tele-rehabilitation, assessing the effectiveness of in-air microphone versus Vocal Holter device, and evaluating different HNR thresholds of 0 dB, -10 dB, and -20 dB (respectively referred to as case A, B, and C). For the first aim, the dataset, which comprises 20 MS patients, is divided into two groups: 10 patients who received tele-rehabilitation and 10 who were treated in-clinic. For each patient, the relevant parameters are reported for three types of vocal material: the sustained vowel /a/, reading, and monologue. In this study, parameters can be extracted using various statistical methods, particularly for monologue and reading; however, the focus is primarily on the mean and median values, as these metrics provide essential insights into the data distribution and allow for a clearer interpretation of the results. The evaluation of rehabilitation methods is conducted separately for the in-air microphone and the VH, utilizing graphical interpretation. For each parameter, the average values of the mean differences between T1 and T0, as well as the standard deviations, are assessed, with attention to the significance of the sign: it is expected that the percentage values of Jitter and Shimmer will decrease in T1 compared to T0, indicating a negative sign; in contrast, HNR, CPPS, and SPL are anticipated to show the opposite trend, resulting in a positive sign. Therefore, the graphs depict a circle representing the average mean values, while the bars above and below are derived from the standard deviations that are multiplied by 2 and divided by the square root of the number of patients in each group, representing the confidence limits. If an overlap between the bar of in-clinc and tele-rehabilitation is presented allows to indicate that the methods are comparable, so the non-inferiority of tele-rehabilitation can be demostrated. During the analysis of the results for each voice material, across both types of microphones, an overlap is observed, indicating that the rehabilitation methods are comparable in all investigated cases. Specifically, in the study of the in-air microphone for reading and monologue, a comparison of different HNR thresholds 0 dB, -10 dB, and -20 dB is conducted to assess various ways of incorporating unharmonic frames into harmonic ones. Using lower thresholds (-10 dB and -20 dB) consistently results in lower values for HNR, CPPS, and SPL, with a greater difference observed between the 0 dB and -10 dB thresholds, while the values between -10 dB and -20 dB remain similar. The dispersion of values with these lower thresholds decreases for SPL and CPPS but increases for HNR. This increase in dispersion for HNR may occur because lower thresholds allow more unharmonic frames to be included, resulting in a wider range of HNR values as these unharmonic components contribute to greater variability in the HNR. Considering the results obtained for each threshold, a graphical representation shows that at T0, all the thresholds allow for comparable values of CPPS, HNR, and SPL; however, at T1, HNR is the only measure that demonstrates a lack of overlap in the confidence interval bars, indicating that the 0 dB threshold produces effects that are not comparable to those of -10 dB and -20 dB. However, despite using these different thresholds, the comparison between in-clinic and tele-rehabilitation remains consistent, demonstrating comparable effects in both techniques. When comparing microphones, it is important to distinguish between vowels, reading, and monologues. In the case of vowels, on average, F0, Shimmer, and CPPS standard deviation show lower values for the VH microphone. This is because F0, and especially Shimmer, are highly sensitive to background and environmental noise, which is captured by the in-air microphone, whereas the VH microphone, which detects vocal fold vibrations, is less affected by noise. The lower CPPS standard deviation in VH reveals that the dispersion of CPPS values is smaller compared to the in-air microphone, indicating a more stable and preferable situation in VH. Meanwhile, CPPS median shows higher values in VH, likely because the VH microphone's low-pass filtering effect reduces interference from higher frequencies, leading to a cleaner signal with more prominent low-frequency components. On the other hand, Jitter shows similar values in both the in-air microphone and VH. In the case of reading and monologue, similar results are obtained, indicating that F0 is higher for the in-air microphone, while SPL is higher for the VH microphone. As explained for vowels, F0 is affected by background noise in the MIC, and higher values can reflect this, as seen in the histograms that show a wider distribution and a bimodal shape for the MIC, indicating the presence of noise. On the other hand, SPL shows higher values for the VH microphone, likely because the VH microphone is in direct contact with the neck, allowing it to capture vocal fold vibrations more efficiently and consistently. This leads to stronger intensity measurements compared to the in-air microphone, which can be more affected by distance and environmental factors. In conclusion, the VH microphone proves to be the better choice for capturing vocal parameters, as it allows for more reliable and consistent evaluations, its ability to minimize the influence of environmental noise, as demonstrated in this study, makes it a more accurate tool for assessing vocal fold vibrations and related acoustic measures. Regarding future developments of this work: it is essential to pay closer attention to the distance between patients and the microphone, as explained, the results at T1 sometimes deviate from the expected theoretical behavior after rehabilitation. This discrepancy may stem from variations in distance or the recording environment; for instance, recordings for T0 and T1 should ideally be conducted in the same room to ensure consistency, additionally it is crucial to ensure that all patients perform the exercises correctly to achieve meaningful improvements. Furthermore, to conduct a more robust analysis, it is necessary to include the results derived from T2, as these results specifically indicate whether the rehabilitation techniques yield improvements over time.It is also important to evaluate the long-term recordings obtained with the VH microphone to assess fatigue, which is particularly significant for patients with multiple sclerosis. In addition, it is interesting to study in more detail whether the use of different HNR thresholds in the pre-processing phase leads to valid outcomes, how these thresholds impact the accuracy and reliability of the measurements obtained during voice analysis, and their significance in the evaluation of patients with multiple sclerosis. Finally, another consideration involves the recordings of reading and monologue with the VH microphone, as they are acquired together. For the file 'Parameters 46 ms', which contains frame-by-frame values for F0 and SPL, it is possible to distinguish between them using a period of silence represented by zeros, so the mean values for both the monologue and reading can be obtained by evaluating these values collectively. However, if the recordings are separated, the 'Parameter Tab' file does not report the average values for the monologue and reading together, instead it provides all parameters with their respective statistics separately for both the monologue and reading. Currently, using the 'Parameters 46 ms' file only yields the average values for F0 and SPL for the monologue and reading separately.

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.

Bibliography

[1] *Comunicare - Definizione*. Online, accessed on July 2024. URL: `https://unaparolaalgiorno.it/significato/comunicare` (cit. on p. 1).

[2] Gustavo Xavier Andrade Miranda. «Analyzing of the vocal fold dynamics using laryngeal videos». PhD thesis. Telecomunicacion, 2017 (cit. on p. 2).

[3] Niv Mor, Kristina Simonyan, and Andrew Blitzer. «Central voice production and pathophysiology of spasmodic dysphonia». In: *The Laryngoscope* 128.1 (2018), pp. 177–183 (cit. on p. 3).

[4] Rhonda R Voskuhl and Stefan M Gold. «Sex-related factors in multiple sclerosis susceptibility and progression». In: *Nature Reviews Neurology* 8.5 (2012), pp. 255–263 (cit. on p. 3).

[5] Alan J Thompson et al. «Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria». In: *The Lancet Neurology* 17.2 (2018), pp. 162–173 (cit. on p. 4).

[6] Massimo Filippi et al. «Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines». In: *Brain* 142.7 (2019), pp. 1858–1875 (cit. on p. 6).

[7] Panagiotis Plotas et al. «Speech deficits in multiple sclerosis: a narrative review of the existing literature». In: *European Journal of Medical Research* 28.1 (2023), p. 252 (cit. on p. 7).

[8] Pamela Seron et al. «Effectiveness of telerehabilitation in physical therapy: a rapid overview». In: *Physical therapy* 101.6 (2021), pzab053 (cit. on p. 7).

[9] *Fondazione Don Carlo Gnocchi*. Online, accessed on August 2024. URL: `https://www.dongnocchi.it/la-fondazione/la-storia/il-nostro-percorso` (cit. on p. 7).

[10] Valeria Crispiatico, Cinzia Baldanzi, Arianna Napoletano, Laura Tomasoni, Francesca Tedeschi, Elisabetta Groppo, Marco Rovaris, Chiara Vitali, and Davide Cattaneo. «Effects of voice rehabilitation in people with MS: A double-blinded long-term randomized controlled trial». In: *Multiple Sclerosis Journal* 28.7 (2022), pp. 1081–1090 (cit. on p. 9).

[11] Chiara Vitali et al. «Telerehabilitation for Lee Silverman Voice Treatment (Tele-LSVT)-Loud on voice intensity and voice use in daily living in people with multiple sclerosis: A protocol for a feasibility and pilot randomized controlled study». In: *Digital Health* 9 (2023), p. 20552076231218150 (cit. on pp. 9, 10, 12).

[12] Marshal F Folstein, Lee N Robins, and John E Helzer. «The mini-mental state examination». In: *Archives of general psychiatry* 40.7 (1983), pp. 812–812 (cit. on p. 10).

[13] «Vocal Holter Med Instruction Manual 2.1.0». In: (2018), pp. 1–24 (cit. on pp. 13, 14).

[14] Antonella Castellana, Alessio Carullo, Simone Corbellini, and Arianna Astolfi. «Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel». In: *IEEE Transactions on Instrumentation and Measurement* 67.3 (2018), pp. 646–654 (cit. on p. 15).

[15] David A Puts, Coren L Apicella, and Rodrigo A Cárdenas. «Masculine voices signal men's threat potential in forager and». In: () (cit. on p. 16).

[16] Antonella Castellana et al. «Towards vocal-behaviour and vocal-health assessment using distributions of acoustic parameters». In: (2018) (cit. on pp. 19, 20).

[17] Paul Boersma et al. «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*. Vol. 17. 1193. Amsterdam. 1993, pp. 97–110 (cit. on p. 21).

[18] João Paulo Teixeira, Carla Oliveira, and Carla Lopes. «Vocal acoustic analysis–jitter, shimmer and hnr parameters». In: *Procedia Technology* 9 (2013), pp. 1112–1122 (cit. on p. 21).

Bibliography

[1] *Comunicare - Definizione*. Online, accessed on July 2024. URL: https://unaparolaalgiorno.it/significato/comunicare (cit. on p. 1).

[2] Gustavo Xavier Andrade Miranda. «Analyzing of the vocal fold dynamics using laryngeal videos». PhD thesis. Telecomunicacion, 2017 (cit. on p. 2).

[3] Niv Mor, Kristina Simonyan, and Andrew Blitzer. «Central voice production and pathophysiology of spasmodic dysphonia». In: *The Laryngoscope* 128.1 (2018), pp. 177–183 (cit. on p. 3).

[4] Rhonda R Voskuhl and Stefan M Gold. «Sex-related factors in multiple sclerosis susceptibility and progression». In: *Nature Reviews Neurology* 8.5 (2012), pp. 255–263 (cit. on p. 3).

[5] Alan J Thompson et al. «Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria». In: *The Lancet Neurology* 17.2 (2018), pp. 162–173 (cit. on p. 4).

[6] Massimo Filippi et al. «Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines». In: *Brain* 142.7 (2019), pp. 1858–1875 (cit. on p. 6).

[7] Panagiotis Plotas et al. «Speech deficits in multiple sclerosis: a narrative review of the existing literature». In: *European Journal of Medical Research* 28.1 (2023), p. 252 (cit. on p. 7).

[8] Pamela Seron et al. «Effectiveness of telerehabilitation in physical therapy: a rapid overview». In: *Physical therapy* 101.6 (2021), pzab053 (cit. on p. 7).

[9] *Fondazione Don Carlo Gnocchi*. Online, accessed on August 2024. URL: https://www.dongnocchi.it/la-fondazione/la-storia/il-nostro-percorso (cit. on p. 7).

[10] Valeria Crispiatico, Cinzia Baldanzi, Arianna Napoletano, Laura Tomasoni, Francesca Tedeschi, Elisabetta Groppo, Marco Rovaris, Chiara Vitali, and Davide Cattaneo. «Effects of voice rehabilitation in people with MS: A double-blinded long-term randomized controlled trial». In: *Multiple Sclerosis Journal* 28.7 (2022), pp. 1081–1090 (cit. on p. 9).

[11] Chiara Vitali et al. «Telerehabilitation for Lee Silverman Voice Treatment (Tele-LSVT)-Loud on voice intensity and voice use in daily living in people with multiple sclerosis: A protocol for a feasibility and pilot randomized controlled study». In: *Digital Health* 9 (2023), p. 20552076231218150 (cit. on pp. 9, 10, 12).

[12] Marshal F Folstein, Lee N Robins, and John E Helzer. «The mini-mental state examination». In: *Archives of general psychiatry* 40.7 (1983), pp. 812–812 (cit. on p. 10).

[13] «Vocal Holter Med Instruction Manual 2.1.0». In: (2018), pp. 1–24 (cit. on pp. 13, 14).

[14] Antonella Castellana, Alessio Carullo, Simone Corbellini, and Arianna Astolfi. «Discriminating pathological voice from healthy voice using cepstral peak prominence smoothed distribution in sustained vowel». In: *IEEE Transactions on Instrumentation and Measurement* 67.3 (2018), pp. 646–654 (cit. on p. 15).

[15] David A Puts, Coren L Apicella, and Rodrigo A Cárdenas. «Masculine voices signal men's threat potential in forager and». In: () (cit. on p. 16).

[16] Antonella Castellana et al. «Towards vocal-behaviour and vocal-health assessment using distributions of acoustic parameters». In: (2018) (cit. on pp. 19, 20).

[17] Paul Boersma et al. «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*. Vol. 17. 1193. Amsterdam. 1993, pp. 97–110 (cit. on p. 21).

[18] João Paulo Teixeira, Carla Oliveira, and Carla Lopes. «Vocal acoustic analysis–jitter, shimmer and hnr parameters». In: *Procedia Technology* 9 (2013), pp. 1112–1122 (cit. on p. 21).